Joint Stocking and Product Offer Decisions Under the Multinomial Logit Model

Huseyin Topaloglu
School of Operations Research and Information Engineering,
Cornell University, Ithaca, New York 14853, USA
Email: topaloglu@orie.cornell.edu
Phone: 607-255-0698 Fax: 607-255-9129

August 27, 2012

Abstract

This paper studies a joint stocking and product offer problem. We have access to a number of products to satisfy the demand over a finite selling horizon. Given that customers choose among the set of offered products according to the multinomial logit model, we need to decide which sets of products to offer over the selling horizon and how many units of each product to stock so as to maximize the expected profit. We formulate the problem as a nonlinear program, where the decision variables correspond to the stocking quantity for each product and the duration of time that each set of products is offered. This nonlinear program is intractable due to its large number of decision variables and its nonseparable and nonconcave objective function. We use the structure of the multinomial logit model to formulate an equivalent nonlinear program, where the number of decision variables is manageable and the objective function is separable. Exploiting separability, we solve the equivalent nonlinear program through a dynamic program with a two dimensional and continuous state variable. Since the solution of the dynamic program requires discretizing the state variable, we study other approximate solution methods. Our equivalent nonlinear program and approximate solution methods yield insights for good offer sets.

Keywords: Product variety; customer choice; assortment; multinomial logit.
History: Received November 2010; revised January 2012; accepted August 2012 by Costis Maglaras after one revision.

In many retail environments, companies offer a set of products that can serve as substitutes of each other and customers make a choice among the set of offered products. Due to substitution possibilities among the products, picking the right set of products to offer and choosing their stocking quantities pose interesting challenges. To begin with, the demand for each product depends on what set of products are actually offered, implying that the demand may not be modeled accurately as an exogenous random variable, as is done in most operational models. Furthermore, the stocking decisions for different products interact with each other so that it may not be appropriate to make the stocking decisions for each product individually. Finally, offering a large set of products ensures that customers are more likely to find what they need, increasing their tendency to make a purchase. However, offering a large set of products also splits the demand among a large number of products, effectively increasing the variability of the demand for each product and resulting in increased safety stocks.

In this paper, we study a joint stocking and product offer problem that tries to address the challenges described above. We have access to a set of products to satisfy the demand from customers that arrive over a finite selling horizon. We need to decide which sets of products to offer and how many units of each product to stock. Over the selling horizon, customers arrive into the system according to a Poisson process and they either make a purchase within the set of offered products or leave without purchasing anything. The choices of customers are governed by the multinomial logit model. The objective is to decide which sets of products to offer over the selling horizon and how many units of each product to stock so as to maximize the expected profit.

We formulate the problem as a nonlinear program, where the decision variables correspond to the stocking quantity for each product and the duration of time that each set of products is offered to customers. There are two difficulties associated with this nonlinear program. First, the number of decision variables is on the order of the number of sets of products, which grows exponentially with the number of products. Second, the objective function is not separable by the products and not necessarily concave. We resolve these difficulties by exploiting the structure of the multinomial logit model. In particular, we show how to formulate an equivalent nonlinear program, where the number of decision variables grows only linearly with the number of products and the objective function is separable by the products. The decision variables in the equivalent nonlinear program correspond to the fractions of customers that choose the different products. Surprisingly, although the decision variables are linked to individual products rather than sets of products, we are still able to capture the substitution possibilities as dictated by the multinomial logit model. Furthermore, since the objective function of the equivalent nonlinear program is separable by the products, we can solve the equivalent nonlinear program by using a dynamic programming formulation with a two dimensional and continuous state variable.

Since the dynamic program that can be used to solve the equivalent nonlinear program has only a two dimensional state variable, we can obtain approximate solutions to it in a tractable fashion by discretizing the state variable. We show that the dynamic program is well behaved in the sense that as we use finer discretizations of the state variable, the loss of precision converges to zero. For a given discretization of the state variable, we give an upper bound on the loss of precision. Despite these positive results for the dynamic program, the necessity to discretize the state space and the lack of intuition into the form of the optimal offer sets motivate us to look for alternative approximate methods for solving our equivalent nonlinear program. Our first approximate method is based on a deterministic approximation to the problem that is formulated under the assumption that the number of customer arrivals takes on its expected value. We give a performance bound for the solution obtained from the deterministic approximation and show that this solution becomes asymptotically optimal as the expected number of customer arrivals gets large. In addition, the solution from the deterministic program always gives priority to products with the largest margins, where margin is computed as the difference between the unit revenue and unit cost. This observation prompts us to work with sets that include a certain number of products with the largest margins. Such sets, noting the asymptotic optimality of the solutions from the deterministic approximation, perform quite well when the expected number of customer arrivals is large, but they may not be satisfactory when working with a small customer volume. Our second approximate method is based on approximating the distribution of the demand for a product with the normal distribution. Our computational experiments demonstrate that the normal approximation perform well over a range of problem parameters. Obtaining solutions under the normal approximation also requires solving a dynamic program with a two dimensional and continuous state variable. To circumnavigate the issues associated with discretizing the state variable of the dynamic program, we finally show how to formulate an integer program that can closely track the solutions to our normal approximation. We bound the gap between the normal approximation and the integer program.

The model in this paper assumes that we can offer different sets of products over the selling horizon and the durations of time over which we offer different sets of products are decision variables. An interesting variant of this problem occurs when we need to choose only one set of products to offer over the whole selling horizon. This variant is significantly more difficult than the one considered in this paper as it is of combinatorial nature. Although the variant with a single offer set is not the main focus of our paper, a welcome feature of the normal approximation described above is that its optimal solution offers a single set over the whole selling horizon. In other words, we can show that one set of products is offered over the whole selling horizon and the other sets are not offered at all in the optimal solution to the normal approximation. The same property also holds for the optimal solution to the integer program approximation described above.

Our paper has strong ties with two streams of work. The model that we study bears resemblance to the one in van Ryzin and Mahajan (1999), where the authors build a joint stocking and product offer model with the stipulation that only one set of products should be offered over the whole selling horizon. Their model essentially corresponds to the variant described in the previous paragraph. In contrast, we allow offering different sets of products over the selling horizon. By doing so, our model avoids the combinatorial aspects of the product offer problem and enables us to extend the work of van Ryzin and Mahajan (1999) in a crucial direction. In particular, to be able to solve their model, the authors in van Ryzin and Mahajan (1999) assume that all of the products have the same unit revenue to cost ratios. This assumption is reasonable in some settings, but it clearly limits the applicability of the model. In this paper, we do not make any assumptions on the cost parameters, extending the work of van Ryzin and Mahajan (1999) to arbitrary cost structures. It is also interesting that although our model allows offering different sets of products over the selling horizon, we can show that if there are n

products under consideration, then it is optimal to offer at most $n+1$ different sets, which is a small fraction of all 2^n possible product sets. Furthermore, these $n+1$ sets are nested in the sense that one of the sets is included in another one, naturally with the exception of the largest set. Therefore, changing the offer sets over time does not require making drastic modifications in the offered assortment.

Our work is also related to the paper by Gallego, Ratliff and Shebalov (2011), where the authors consider linear programming formulations for network revenue management problems under customer choice behavior. These linear programs have one decision variable for each set of itineraries, corresponding to the number of time periods over which a set of itineraries remains open for purchase. The authors show how to formulate an equivalent linear program with fewer decision variables when the choices of customers follow the multinomial logit model. The equivalent nonlinear program that we give in this paper is initiated by their work. Gallego et al. (2011) use linear programming duality to show the equivalence between their linear programs. In this paper, we deal with nonlinear programs, but we are able to use relationships between primal solutions, avoiding linear programming duality. Therefore, our paper extends the work done by Gallego et al. (2011) to nonlinear programs. It is also important to note that our equivalent nonlinear program not only has smaller number of decision variables than the original one, but it also has a separable objective function. The separability allows us to pursue algorithmic approaches based on dynamic or integer programming, which are not possible for the original nonlinear program. Ultimately, our paper and the work done by Gallego et al. (2011) collectively point out that the multinomial logit model possesses a special structure that can be exploited to build tractable operational models. Our hope is that other researchers pursue the results in this paper and in Gallego et al. (2011) in new settings.

To sum up, we make the following research contributions in this paper. 1) Building on van Ryzin and Mahajan (1999), we formulate a joint stocking and product offer model. An important differentiating aspect of our model is that it allows products with arbitrary unit revenues and costs. A naive formulation of our model has a large number of decision variables and a nonseparable objective function, but we give an equivalent formulation with a small number of decision variables and a separable objective function. 2) We show how to exploit the separability of the objective function to solve our model by using a dynamic program with a two dimensional and continuous state variable. The numerical solution of the dynamic program requires discretizing the state variable. We bound the loss in precision due to discretization. 3) We show that if there are n products under consideration, then the optimal solution of our model offers at most $n+1$ different sets of products over the selling horizon. 4) We study approximate solution methods for our model. The first approximate method is based on a deterministic approximation. We bound the performance of the solutions obtained from the deterministic approximation. Furthermore, this approximation method motivates using sets that include a certain number of products with the largest margins. 5) We investigate a second approximate solution method that is based on approximating the demand distributions with the normal distribution. The solution to the normal approximation requires solving a dynamic program, but we formulate an integer program that closely tracks the solution to the normal approximation. We establish that the solutions from the normal and integer programming approximations naturally offer a single set over the whole selling horizon, even though we do not explicitly impose this requirement. 6) We provide computational experiments that test the performance of our approximation methods and demonstrate that the normal approximation and the associated integer programming formulation perform remarkably well.

The paper is organized as follows. Section 1 reviews the related literature. Section 2 formulates our joint stocking and product offer model. Section 3 transforms the model into a more tractable equivalent formulation. Section 4 uses a dynamic program to solve the equivalent formulation. The next three sections focus on approximations of the equivalent formulation. Section 5 focuses on a deterministic approximation, Section 6 provides the normal approximation and Section 7 gives an approximation by using an integer program. Section 8 presents computational experiments. Section 9 concludes.

1 Literature Review

It is customary to differentiate between static and dynamic substitution when building joint stocking and product offer models. In static substitution, customers make a choice among the set of offered products without observing the availability of stock. If a customer chooses a product for which no stock is available, then the customer does not attempt to make a second choice. In other words, the choices of customers are influenced by the set of offered products, but not by the availability of stock. Static substitution is applicable when customers make a choice from a catalogue or a floor display. The model in van Ryzin and Mahajan (1999) and our model in this paper focus on static substitution under the multinomial logit model. Smith and Agrawal (2000) study a static substitution model that can penalize and limit the number of offered products. In dynamic substitution, customers observe the availability of stock and make a choice only among the products for which stocks are available. Mahajan and van Ryzin (2001) use stochastic approximation algorithms to compute stocking levels under dynamic substitution. Gaur and Honhon (2006) analyze a product offer problem under the choice model of Lancaster (1966) and build on their results for static substitution to suggest heuristics for dynamic substitution. Kok and Fisher (2007) study dynamic substitution under the multinomial logit model and propose methods for estimating choice model parameters and making stocking decisions. Honhon, Gaur and Seshadri (2010) analyze a dynamic substitution model, where customers are classified into types depending on their orders of preference for different products. Goyal, Levi and Segev (2009) develop approximation algorithms under a fixed preference order and dynamic substitution.

Customer choice is an active area of research in revenue management. In revenue management, it is common to assume that stocking decisions are fixed, as in the case for fixed seating capacities on flight legs. Assuming that customers choose between the offered itineraries according to a particular choice model, the focus is on how to dynamically adjust the set of offered itineraries as a function of the remaining seat inventory and time left in the selling horizon. Talluri and van Ryzin (2004) study revenue management problems over a single flight leg and characterize the structure of the optimal policy. Gallego, Iyengar, Phillips and Dubey (2004) formulate a linear programming approximation for revenue management problems with customer choice behavior over a network of flight legs. Liu and van Ryzin (2008) give a dynamic programming formulation for network revenue management problems and approximately solve the dynamic program by decomposing it by the flight legs. The paper by van Ryzin and Vulcano (2008) uses stochastic approximation algorithms to compute protection level policies. For the linear program proposed by Gallego et al. (2004), Bront, Diaz and Vulcano (2009) consider the case where there are multiple customer types that make a choice according to multinomial logit models with different parameters. Gallego et al. (2011) focus on the same linear program and show how to reduce the number of decision variables from exponential in the number of possible itineraries to linear when there is a single customer type making a choice according to the multinomial logit model. Zhang and Cooper (2005) study revenue management problems where customers choose between parallel flight legs between a fixed origin destination pair. They point out that the problem is computationally difficult and give bounds on the optimal expected revenue. If the stocking quantities are fixed and we are interested in finding a policy to dynamically adjust the set of offered products as a function of the remaining inventories and the remaining time in the selling horizon, then the problem in this paper and the problem in Zhang and Cooper (2005) are equivalent. Thus, even if the stocking quantities are fixed in our problem, finding the optimal policy to dynamically adjust the offered products is difficult.

The ideas in this paper may help develop more intricate network revenue management models. Under the assumption that customers arrive into the system with the intention of purchasing a fixed itinerary, there are two approaches for building approximations to network revenue management problems. Using D_i to denote the total demand for itinerary i, the first approach, known as the deterministic linear program, assumes that the demand for itinerary i is equal to the expected value of D_i. The decision variable x_i in the deterministic linear program corresponds to the number of tickets sold for itinerary i, satisfying the constraint $x_i \leq \mathbb{E}\{D_i\}$. The second approach, known as the probabilistic nonlinear program, uses the number of tickets made available for itinerary i as the decision variable. If we make y_i tickets available for itinerary i, then the expected sales for itinerary i is $\mathbb{E}\{\min\{y_i, D_i\}\}$, which is nonlinear in y_i. We refer the reader to Talluri and van Ryzin (2005) for deterministic linear and probabilistic nonlinear programs. For the case where customers choose between the offered itineraries, the linear programming approximations in Gallego et al. (2004) and Liu and van Ryzin (2008) are analogues of the deterministic linear program. This paper shows that solving nonlinear optimization problems under the multinomial logit model is tractable. Thus, our approach may allow formulating analogues of the probabilistic nonlinear program under customer choice behavior.

2 Problem Formulation

Customers arrive according to a Poisson process with rate λ. For brevity, we normalize the length of the selling horizon to one. There are n products indexed by $1, \ldots, n$. The unit purchasing cost and revenue of product i are respectively c_i and p_i with $p_i \geq c_i$. We let x_i be the number of units of product i that we stock at the beginning of the selling horizon. Since customers choose among the set of offered products, we can influence the demand by adjusting the set of products that we offer over the selling horizon. If we offer the set of products $S \subseteq \{1, \ldots, n\}$, then the probability that a customer chooses product i is denoted by $P_i(S)$. We assume that the choice probability $P_i(S)$ follows the multinomial logit model. Under the multinomial logit model, each customer associates the preference weight v_i with product i and the preference weight v_0 with the no purchase option. We use $V(S)$ to denote the total preference weight of the available options when we offer the set S so that $V(S) = v_0 + \sum_{i \in S} v_i$. According to the multinomial logit model, if we offer the set of products S, then each customer chooses product

$i \in S$ with probability $P_i(S) = v_i/V(S)$; see McFadden (1980). We have $P_i(S) = 0$ when $i \notin S$. With the remaining probability $P_0(S) = 1 - \sum_{i=1}^n P_i(S) = v_0/V(S)$, each customer leaves without purchasing anything. We use a static policy to adjust the set of offered products over the selling horizon, where we offer the set of products S for $y(S)$ time units. The policy is static in the sense that we choose the values of the decision variables $\{y(S) : S \subseteq \{1, \ldots, n\}\}$ at the beginning of the selling horizon.

We offer the set of products S during a time period of length $y(S)$ time units. The number of customer arrivals during this time period has a Poisson distribution with mean $\lambda\, y(S)$. Furthermore, each customer that arrives during this time period chooses product i with probability $P_i(S)$. In this case, if we assume that the choice of each customer is independent of the others, then the demand for product i during the time period that we offer the set of products S has a Poisson distribution with mean $\lambda\, P_i(S)\, y(S)$. Using $\mathsf{Pois}(\alpha)$ to denote a Poisson random variable with mean α, we can maximize the expected profit by solving the problem

$$\max \quad \sum_{i=1}^n p_i\, \mathbb{E}\Big\{ \min \Big\{ \mathsf{Pois}\big(\textstyle\sum_{S \subseteq \{1,\ldots,n\}} \lambda\, P_i(S)\, y(S) \big), x_i \Big\} \Big\} - \displaystyle\sum_{i=1}^n c_i\, x_i \tag{1}$$

$$\text{subject to} \quad \sum_{S \subseteq \{1,\ldots,n\}} y(S) = 1 \tag{2}$$

$$y(S), x_i \geq 0 \qquad S \subseteq \{1, \ldots, n\},\ i = 1, \ldots, n, \tag{3}$$

where the first term in the objective function is the expected revenue from the choices of customers and the second term is the stocking cost of the products. There are several difficulties associated with the problem above. First, the number of decision variables is $n + 2^n$, growing exponentially with the number of products. Second, it is possible to show that the objective function is concave in the decision variables $y = \{y(S) : S \subseteq \{1, \ldots, n\}\}$ and $x = (x_1, \ldots, x_n)$ separately, but it is not necessarily jointly concave in (y, x). Third, the decisions for the different products interact with each other through the decision variables y and we cannot concentrate on each product individually.

In the next section, we give a tractable solution approach for problem (1)-(3). Before doing so, however, it is useful to briefly elaborate on the implications of using a static policy to decide how to adjust the set of products that we offer for purchase. As we offer the set of products S for $y(S)$ units of time, we may run out of stock for a product in the set S. The implicit assumption in problem (1)-(3) is that if this situation happens to be the case, then we continue offering the set of products S and when a customer chooses a product in the set S for which we do not have any stock, the customer simply leaves the system without purchasing anything. This is the standard static substitution assumption that we mention in Section 1 and the model in van Ryzin and Mahajan (1999) makes the same assumption. This assumption can be justified in several ways. If the customer chooses a product for which we do not have any stock, then we may replenish the stock with an emergency procurement at a cost and it is simple to incorporate the cost of such an emergency procurement into our model. In this case, van Ryzin and Mahajan (1999) interpret the set of offered products as a floor display, among which customers make a choice without paying attention to the stock. Another option is to solve our model once again whenever we run out of stock for a certain product and recompute a new static policy to decide how to adjust the set of offered products over the remaining portion of the selling horizon.

3 Equivalent Formulation

In this section, we give an equivalent formulation for problem (1)-(3) whose objective function decomposes by the products, in which case, we can solve problem (1)-(3) by using a dynamic program. In the equivalent formulation, we let w_i be the fraction of customers that choose product i and w_0 be the fraction of customers that choose to leave without purchasing anything. Naturally, these decision variables have to satisfy the constraint $\sum_{i=1}^n w_i + w_0 = 1$. During the time period that we offer the set of products S, the fraction of customers that choose product i is $P_i(S) = v_i/V(S)$ whenever $i \in S$ and zero otherwise. On the other hand, the fraction of customers that choose to leave without purchasing anything is $P_0(S) = v_0/V(S)$. Therefore, we always have $P_i(S)/v_i \le P_0(S)/v_0$ for all $S \subseteq \{1,\ldots,n\}$, $i = 1,\ldots,n$ and it is reasonable to impose the constraint $w_i/v_i \le w_0/v_0$ for all $i = 1,\ldots,n$. In this case, we propose maximizing the expected profit by solving the problem

$$\max \quad \sum_{i=1}^n p_i\,\mathbb{E}\big\{\min\big\{\mathsf{Pois}(\lambda\, w_i), x_i\big\}\big\} - \sum_{i=1}^n c_i\, x_i \tag{4}$$

$$\text{subject to} \quad \sum_{i=1}^n w_i + w_0 = 1 \tag{5}$$

$$\frac{w_i}{v_i} - \frac{w_0}{v_0} \le 0 \qquad i = 1,\ldots,n \tag{6}$$

$$w_i, w_0, x_i \ge 0 \qquad i = 1,\ldots,n. \tag{7}$$

We observe that the objective function of the problem above decomposes by the products and there are $2n+1$ decision variables in this problem, as opposed to the $n + 2^n$ decision variables in problem (1)-(3). Furthermore, if we fix the value of the decision variable w_0, then the decision variables $w = (w_1,\ldots,w_n)$ interact only through constraint (5). Constraints (6) are the only part where the multinomial logit model plays a role and it turns out that these constraints are adequate to capture the choice behavior of customers. The next theorem shows that problems (1)-(3) and (4)-(7) are equivalent to each other. We note that this equivalence is dependent on the structure of the multinomial logit model and it does not necessarily hold under other choice models.

Theorem 1 *Problems* (1)-(3) *and* (4)-(7) *are equivalent to each other. In other words, given an optimal solution to one problem, we can construct a feasible solution to the other one that provides the same objective value.*

Proof. We give the main idea of the proof, but defer the details to Appendix A.1. First, we assume that (y^*, x^*) is an optimal solution to problem (1)-(3) providing the objective value Z^*. We define the solution (w^*, w_0^*, x^*) to problem (4)-(7) by letting

$$w_0^* = \sum_{S\subseteq\{1,\ldots,n\}} P_0(S)\, y^*(S) \quad \text{and} \quad w_i^* = \sum_{S\subseteq\{1,\ldots,n\}} P_i(S)\, y^*(S) \tag{8}$$

for all $i = 1,\ldots,n$. In this case, we verify in Appendix A.1 that the solution (w^*, w_0^*, x^*) satisfies all of the constraints in problem (4)-(7) and provides the objective value Z^*. Second, we assume that

(w^*, w_0^*, x^*) is an optimal solution to problem (4)-(7) providing the objective value Z^*. Without loss of generality, we order the products so that $w_1^*/v_1 \geq w_2^*/v_2 \geq \ldots \geq w_n^*/v_n$. Noting constraints (6), we also have $w_0^*/v_0 \geq w_1^*/v_1$. We label the sets $S_0 = \varnothing$ and $S_i = \{1, \ldots, i\}$ for all $i = 1, \ldots, n$. In this case, we define the solution (y^*, x^*) to problem (1)-(3) by letting

$$y^*(S_i) = \left[\frac{w_i^*}{v_i} - \frac{w_{i+1}^*}{v_{i+1}} \right] V(S_i) \tag{9}$$

for all $i = 0, \ldots, n$ with the convention that $y^*(S_n) = [w_n^*/v_n]\, V(S_n)$. We set $y^*(S) = 0$ whenever $S \notin \{S_0, S_1, \ldots, S_n\}$. In Appendix A.1, we verify that the solution (y^*, x^*) satisfies all of the constraints in problem (1)-(3) and provides the objective value Z^*. □

The proof of Theorem 1 shows how to use problem (4)-(7) and the transformation in (9) to obtain an optimal solution to problem (1)-(3). The decision variables in problem (4)-(7) represent fractions and have nothing to do with the choices of customers, but constraints (6) are adequate to capture the choices stipulated by the multinomial logit model. Gallego et al. (2011) establish a similar result by considering a linear programming formulation for network revenue management problems under customer choice behavior. Assuming that the choices of customers are governed by the multinomial logit model, the authors transform this linear program into an equivalent linear program with fewer decision variables. Their proof of equivalence for the two linear programs uses linear programming duality. The objective functions of problems (1)-(3) and (4)-(7) are nonlinear and not necessarily concave, prohibiting the use of duality. Therefore, we work with primal solutions in the proof of Theorem 1.

To gain more insight into the equivalence between problems (1)-(3) and (4)-(7), we note that if we offer the set of products S, then under the multinomial logit model, the fraction $P_i(S)$ of customers that choose product i satisfies $P_i(S)/v_i = 1/V(S) = P_0(S)/v_0$ for all $i \in S$, and $P_i(S)/v_i = 0 \leq P_0(S)/v_0$ for all $i \notin S$. In the last two chains of equalities and inequalities, we use the fact that $P_i(S) = v_i/V(S)$ when $i \in S$ and $P_i(S) = 0$ when $i \notin S$. Therefore, we always have $P_i(S)/v_i \leq P_0(S)/v_0$ and it is sensible that the fraction of customers that choose product i and the fraction of customers that leave without choosing anything satisfy constraints (6). Similarly, the fractions of customers that choose a product and that leave without choosing anything should add up to one, justifying constraint (5). Thus, if $(w_1, \ldots, w_n)$ are the fractions of customers that choose the different products under the multinomial logit model and w_0 is the fraction of customers that leave without choosing anything, then (w, w_0) must satisfy the constraints in problem (1)-(3).

Conversely, if we consider (w, w_0) that satisfies the constraints in problem (1)-(3), then it must be the case that w_i corresponds to a fraction of customers that choose product i under the multinomial logit model. To see that this is indeed the case, the critical observation is that if $(\tilde{w}, \tilde{w}_0)$ is an extreme point of the set of feasible solutions to problem (4)-(7), then there exists a set of products $\tilde{S}$ associated with this extreme point such that $\tilde{w}_i = P_i(\tilde{S})$ for all $i = 0, \ldots, n$. This observation follows from an immediate corollary to the proof of Theorem 1 and we provide the corollary in Appendix A.2. So, any extreme point of the set of feasible solutions to problem (4)-(7) is associated with a set of products. On the other hand, any feasible solution (w, w_0) to problem (4)-(7) can be written as $(w, w_0) = \sum_{k=1}^K \alpha^k\, (\tilde{w}^k, \tilde{w}_0^k)$, where $(\tilde{w}^k, \tilde{w}_0^k)$ is an extreme point and we have $\sum_{k=1}^K \alpha^k = 1$ with $\alpha^k \geq 0$ for all $k = 1, \ldots, K$. Using $\tilde{S}^k$

to denote the set of products associated with the extreme point $(\tilde{w}^k, \tilde{w}_0^k)$ and using the fact that $\tilde{w}_i^k = P_i(\tilde{S}^k)$ as discussed at the beginning of this paragraph, the last equality can be written componentwise as $w_i = \sum_{k=1}^K \alpha^k P_i(\tilde{S}^k)$ for all $i = 0, \ldots, n$. Thus, if we consider (w, w_0) that is feasible to problem (4)-(7), then w_i should be the fraction of customers that choose product i under the multinomial logit model when we offer the set $\tilde{S}^k$ associated with the extreme point $(\tilde{w}^k, \tilde{w}_0^k)$ with frequency α^k. To sum up, the discussion in the previous paragraph indicates that if (w, w_0) gives the fractions of customers that choose the different products and that leave without choosing anything under the multinomial logit model, then (w, w_0) must satisfy the constraints in problem (4)-(7). Conversely, the discussion in this paragraph indicates that if (w, w_0) satisfies the constraints in problem (4)-(7), then w_i corresponds to a fraction of customers that choose product i under the multinomial logit model. These two statements ensure that problems (1)-(3) and (4)-(7) are equivalent to each other.

We can use Theorem 1 to show that we actually do not offer too many different sets of products when we follow the static policy suggested by the optimal solution to problem (1)-(3). In particular, the next proposition shows that if we follow the static policy suggested by the optimal solution to problem (1)-(3), then we offer at most $n + 1$ different sets and these sets are nested in the sense that each one of the sets is included in another one, with the exception of the largest one of them.

Proposition 2 *There exists an optimal solution* (y^*, x^*) *to problem* (1)-(3) *with at most* $n + 1$ *of the decision variables* y^* *taking nonzero values. Furthermore, if we let* $y^*(S_0), y^*(S_1), \ldots, y^*(S_n)$ *be the* $n+1$ *of the decision variables* y^* *that can take nonzero values, then we have* $S_0 \subseteq S_1 \subseteq \ldots \subseteq S_n$.

Proof. The proof follows from the second half of the proof of Theorem 1. If we let (w^*, w_0^*, x^*) be an optimal solution to problem (4)-(7), define the sets $S_0, S_1, \ldots, S_n$ as in the proof of Theorem 1 and construct the solution (y^*, x^*) as in (9), then it follows from Theorem 1 that (y^*, x^*) is an optimal solution to problem (1)-(3). Furthermore, the definition of S_i implies that we have $S_0 \subseteq S_1 \subseteq \ldots \subseteq S_n$ and only $y^*(S_0), y^*(S_1), \ldots, y^*(S_n)$ among the decision variables y^* can take nonzero values. □

Proposition 2 provides a practically appealing approach for implementing the static policy that we obtain from problem (1)-(3). In particular, we can start by offering the largest set of products S_n for $y^*(S_n)$ time units and switch to the next smaller set as time progresses. Each switch from the current set to the next smaller one requires taking one product out of the current set, avoiding the need to drastically change the set of products that we offer.

4 Solution Approach

Problem (4)-(7) has fewer decision variables than problem (1)-(3), but this problem can still be difficult to solve directly since its objective function is not necessarily jointly concave in (w, w_0, x). To overcome this difficulty, we observe that the objective function of problem (4)-(7) decomposes by the products. Furthermore, if we fix the value of the decision variable w_0, then problem (4)-(7) has a knapsack structure. The decision variable w_i represents the number of units of product i that we put into the knapsack. Noting constraint (5), the capacity of the knapsack is $1 - w_0$ and we have to fill the knapsack exactly up to its capacity. Constraints (6) imply that we cannot put more than $v_i\, w_0/v_0$ units of product i into the knapsack. If we put w_i units of product i, then we generate a utility of $p_i\, \mathbb{E}\{\min\{\mathsf{Pois}(\lambda\, w_i), x_i\}\} - c_i\, x_i$, where the decision variable x_i is chosen to maximize this utility from product i. Viewing problem (4)-(7) as a knapsack problem allows us to use a dynamic program to solve this problem for a fixed value of w_0. The decision epochs correspond to the products. The action variables in decision epoch i are (w_i, x_i). We use z_i to denote the state variable in decision epoch i, corresponding to the total capacity used by the earlier products. Therefore, we can obtain the optimal objective value of problem (4)-(7) for a fixed value of w_0 by solving the optimality equation

$$\begin{aligned} \Theta_i(z_i \,|\, w_0) = \max \quad & p_i\, \mathbb{E}\big\{\min\big\{\mathsf{Pois}(\lambda\, w_i), x_i\big\}\big\} - c_i\, x_i + \Theta_{i+1}(z_i + w_i \,|\, w_0) && (10)\\ \text{subject to} \quad & w_i \le 1 - w_0 - z_i && (11)\\ & \frac{w_i}{v_i} \le \frac{w_0}{v_0} && (12)\\ & w_i, x_i \ge 0, && (13) \end{aligned}$$

where the boundary condition is $\Theta_{n+1}(z_{n+1} \,|\, w_0) = 0$ if $z_{n+1} = 1 - w_0$ and $\Theta_{n+1}(z_{n+1} \,|\, w_0) = -\infty$ if $z_{n+1} < 1 - w_0$. This boundary condition ensures that it is optimal to terminate at the states that consume all of the capacity of the knapsack. In this case, we can solve the problem $\max_{w_0 \in [0,1]} \Theta_1(0 \,|\, w_0)$ to obtain the optimal objective value of problem (4)-(7). The state variable in the optimality equation in (10)-(13) appears to be single dimensional, but since this dynamic program needs to be solved for all values of w_0, its state variable is effectively (z_i, w_0). We observe that it is straightforward to find the best value of the action variable x_i corresponding to each value of w_i by solving the problem $\max_{x_i \ge 0} p_i\, \mathbb{E}\{\min\{\mathsf{Pois}(\lambda\, w_i), x_i\}\} - c_i\, x_i$. The last problem is a newsvendor problem and its optimal solution is given by the smallest x_i that satisfies $\mathbb{P}\{\mathsf{Pois}(\lambda\, w_i) \le x_i\} \ge 1 - c_i/p_i$. Therefore, the only crucial action variable in the optimality equation is w_i.

In the optimality equation in (10)-(13), the state variable (z_i, w_0) and the action variable w_i are continuous. Thus, the numerical solution of this optimality equation requires discretizing the state and action variables, in which case, it is natural to ask about the loss of precision when we solve the discretized version of the optimality equation. To answer this question, we begin by studying problem (4)-(7) when the decision variables in this problem are discretized. In particular, we let $X_i(w_i)$ be the smallest x_i that satisfies $\mathbb{P}\{\mathsf{Pois}(\lambda\, w_i) \le x_i\} \ge 1 - c_i/p_i$, in which case, $X_i(w_i)$ is the best value of the decision variable x_i in problem (4)-(7) for a given value of w_i. Noting the objective function of problem (4)-(7), we define $G_i(w_i) = p_i\, \mathbb{E}\{\min\{\mathsf{Pois}(\lambda\, w_i), X_i(w_i)\}\} - c_i\, X_i(w_i)$ so that problem (4)-(7) is equivalent to maximizing $\sum_{i=1}^n G_i(w_i)$ over $(w, w_0) \ge 0$ subject to constraints (5) and (6). In the next proposition, we bound the derivative of $G_i(\cdot)$. The proof is deferred to Appendix A.3.

Proposition 3 *The directional derivative of $G_i(\cdot)$ at any point is bounded by $(p_i - c_i)\, \lambda$.*

Proposition 3 allows us to bound the loss of precision when we solve problem (4)-(7) after discretizing the decision variables. To see this, assume that the optimal solution to the original version of problem (4)-(7) is $(w^*, w_0^*, X(w^*))$, where we let $X(w) = (X_1(w_1), \ldots, X_n(w_n))$. On the other hand, letting L

be an integer, we discretize the interval $[0,1]$ by using the set of points $\mathcal{L} = \{k/L : k = 0, 1, \ldots, L\}$ and consider the discretized version of problem (4)-(7) where each one of the decision variables (w, w_0) is restricted to take a value in the set $\mathcal{L}$. Since the successive points in $\mathcal{L}$ are separated by $1/L$, we can obtain a solution $(\hat{w}, \hat{w}_0, X(\hat{w}))$ to the discretized version of problem (4)-(7) by letting

$$\hat{w}_i = \frac{1}{L} \lfloor L\, w_i^* \rfloor \qquad i = 1, \ldots, n \quad \text{and} \quad \hat{w}_0 = 1 - \sum_{i=1}^{n} \hat{w}_i, \tag{14}$$

where we use $\lfloor \cdot \rfloor$ to denote the round down function. We can check that the solution $(\hat{w}, \hat{w}_0, X(\hat{w}))$ is feasible to the discretized version of problem (4)-(7). In particular, we have $\hat{w}_i \in \mathcal{L}$ for all $i = 0, \ldots, n$ by the construction in (14). Similarly, by construction, $\sum_{i=1}^n \hat{w}_i + \hat{w}_0 = 1$ ensuring that the solution $(\hat{w}, \hat{w}_0, X(\hat{w}))$ satisfies constraint (5) in problem (4)-(7). Also, we have $\hat{w}_i = \frac{1}{L} \lfloor L\, w_i^* \rfloor \leq w_i^*$ for all $i = 1, \ldots, n$, in which case, noting that $(w^*, w_0^*, X(w^*))$ is feasible to problem (4)-(7), it follows that $w_0^* = 1 - \sum_{i=1}^n w_i^* \leq 1 - \sum_{i=1}^n \hat{w}_i = \hat{w}_0$. Once more, using the fact that $(w^*, w_0^*, X(w^*))$ is feasible to problem (4)-(7) and noting that $\hat{w}_i \leq w_i^*$ for all $i = 1, \ldots, n$ and $w_0^* \leq \hat{w}_0$, we obtain $\hat{w}_i / v_i \leq w_i^* / v_i \leq w_0^* / v_0 \leq \hat{w}_0 / v_0$ for all $i = 1, \ldots, n$, establishing that the solution $(\hat{w}, \hat{w}_0, X(\hat{w}))$ satisfies constraints (6) in problem (4)-(7) as well. Thus, $(\hat{w}, \hat{w}_0, X(\hat{w}))$ is feasible to the discretized version of problem (4)-(7). To bound the loss of precision by using the solution $(\hat{w}, \hat{w}_0, X(\hat{w}))$, we note that

$$|w_i^* - \hat{w}_i| = \left| \frac{1}{L} L\, w_i^* - \frac{1}{L} \lfloor L\, w_i^* \rfloor \right| \leq \frac{1}{L}$$

by (14). In this case, Proposition 3 implies that $\sum_{i=1}^n G_i(w_i^*) - \sum_{i=1}^n G_i(\hat{w}_i) \leq \frac{1}{L} \sum_{i=1}^n (p_i - c_i)\, \lambda$. Since $\sum_{i=1}^n G_i(w_i^*)$ is the optimal objective value of problem (4)-(7), the last inequality shows that if we solve the discretized version of problem (4)-(7) instead of the original version, then the loss of precision is no larger than $\frac{1}{L} \sum_{i=1}^n (p_i - c_i)\, \lambda$. The loss of precision converges to zero as L goes to infinity.

The discussion in the previous paragraph bounds the loss of precision when we solve a discretized version of problem (4)-(7), where each one of the decision variables (w, w_0) is restricted to take a value in $\mathcal{L}$. On the other hand, we observe that the solution to this discretized version of problem (4)-(7) can be obtained precisely by solving the optimality equation in (10)-(13) with the state variable (z_i, w_0) taking values in $\mathcal{L}^2$ and the action variable w_i taking values in $\mathcal{L}$. Therefore, by solving a dynamic program with n decision epochs, $(L+1)^2$ possible values for the state variable at each decision epoch and $L + 1$ possible values for the action variable at each state, we can obtain a solution to problem (4)-(7) and the objective value provided by this solution would deviate from the optimal objective value of problem (4)-(7) by at most an additive factor of $\frac{1}{L} \sum_{i=1}^n (p_i - c_i)\, \lambda$.

5 Deterministic Approximation

In this section, we study a deterministic approximation to our model that is formulated under the assumption that the number of customer arrivals takes on its expected value. We have two goals when studying this deterministic approximation. First, we would like to bound the performance of the solutions that we obtain from the deterministic approximation. In this way, we can characterize the situations where we can expect such an approximation to work well. Second, by using the solution to the deterministic approximation, we would like to characterize structural properties of good offer sets. The deterministic approximation we consider in this section is obtained by replacing the random variable $\mathsf{Pois}\big(\sum_{S\subseteq\{1,\ldots,n\}} \lambda\, P_i(S)\, y(S)\big)$ in the objective function of problem (1)-(3) with its expectation $\sum_{S\subseteq\{1,\ldots,n\}} \lambda\, P_i(S)\, y(S)$. In this case, we can follow precisely the same line of reasoning followed in Section 3 to show that this deterministic version of problem (1)-(3) is equivalent to maximizing $\sum_{i=1}^n p_i \min\{\lambda\, w_i, x_i\} - \sum_{i=1}^n c_i\, x_i$ over $(w, w_0, x) \geq 0$ subject to constraints (5) and (6). Furthermore, since $p_i \geq c_i$, the optimal value of the decision variable x_i in the last problem is always $\lambda\, w_i$, in which case, the objective function of the last problem becomes $\sum_{i=1}^n (p_i - c_i)\, \lambda\, w_i$. Thus, the deterministic version of problem (1)-(3) that we study in this section is equivalent to the problem

$$\begin{aligned} \max \quad & \sum_{i=1}^n (p_i - c_i)\, \lambda\, w_i && && (15)\\ \text{subject to} \quad & \sum_{i=1}^n w_i + w_0 = 1 && && (16)\\ & \frac{w_i}{v_i} - \frac{w_0}{v_0} \leq 0 && i = 1, \ldots, n && (17)\\ & w_i, w_0 \geq 0 && i = 1, \ldots, n. && (18) \end{aligned}$$

By the same argument in the proof of Theorem 1, we can show that if (w^*, w_0^*) is an optimal solution to problem (15)-(18), then defining the solution (y^*, x^*) such that $x_i^* = \lambda\, w_i^*$ and y^* is as in (9), (y^*, x^*) is an optimal solution to the deterministic version of problem (1)-(3).

In the next proposition, we bound the performance of the solution that we obtain from problem (15)-(18). To this end, we let $Q_i(w_i, x_i) = p_i\, \mathbb{E}\{\min\{\mathsf{Pois}(\lambda\, w_i), x_i\}\} - c_i\, x_i$ so that the objective function of problem (4)-(7) is $\sum_{i=1}^n Q_i(w_i, x_i)$. We let $(w^*, w_0^*, X(w^*))$ be an optimal solution to problem (4)-(7) and (w^D, w_0^D) be an optimal solution to problem (15)-(18), where we use $X(w)$ as defined in the previous section. We have $\sum_{i=1}^n Q_i(w_i^D, X_i(w_i^D)) / \sum_{i=1}^n Q_i(w_i^*, X_i(w_i^*)) \leq 1$ since $(w^D, w_0^D, X(w^D))$ is a feasible but not necessarily an optimal solution to problem (4)-(7). The next proposition gives a lower bound on $\sum_{i=1}^n Q_i(w_i^D, X_i(w_i^D)) / \sum_{i=1}^n Q_i(w_i^*, X_i(w_i^*))$, bounding the relative loss in the expected profit by using the solution $(w^D, w_0^D, X(w^D))$ in problem (4)-(7) instead of the optimal solution $(w^*, w_0^*, X(w^*))$. We defer the proof of this proposition to Appendix A.4. In this proposition and throughout the rest of this section, we assume that the products are indexed according to their margins so that we have $p_1 - c_1 > p_2 - c_2 > \ldots > p_n - c_n$.

Proposition 4 *It holds that*

$$\frac{\sum_{i=1}^n Q_i(w_i^D, X_i(w_i^D))}{\sum_{i=1}^n Q_i(w_i^*, X_i(w_i^*))} \geq 1 - \frac{\sqrt{\sum_{i=1}^n p_i^2}}{2\sqrt{\lambda}\, (p_1 - c_1)\, \frac{v_1}{v_0 + v_1}}. \tag{19}$$

In Proposition 4, as the arrival rate λ increases or the preference weight of the no purchase option v_0 decreases, the right side of (19) increases. When λ is large, the expected number of customer arrivals is large and when v_0 is small, a large fraction of customers actually make a choice within the offered set of products. Thus, Proposition 4 indicates that the solution obtained from the deterministic approximation should perform better when the expected number of customers who make a choice within the offered set of products is larger. Also, as $p_1 - c_1$ gets larger, the right side of (19) increases as well, indicating that the solution obtained from the deterministic approximation may not perform well when the margins of the products are too small. Our computational experiments confirm these expectations. To get a feel for the performance bound in Proposition 4, consider the case where the unit revenues are in the range $[0, \delta]$ and the largest margin product is the most expensive one with a relative margin of 50% so that $p_1 - c_1 = \delta/2$. For simplicity, assume that all customers choose a product so that $v_0 = 0$. In this case, the right side of (19) exceeds $1 - \sqrt{n/\lambda}$. This is to say that the solution from the deterministic approximation obtains at least $100\,(1 - \sqrt{n/\lambda})\%$ of the optimal expected profit.

Performance bounds of the form in (19) are common in the revenue management literature, dating back to the work of Gallego and van Ryzin (1994) and Gallego and van Ryzin (1997), where the authors give performance bounds for policies obtained from deterministic approximations of dynamic pricing problems. Talluri and van Ryzin (1998) extend this work to network revenue management problems. For revenue management problems with reusable resources, Levi and Radovanovic (2010) give performance bounds for policies obtained from deterministic approximations. Perhaps more interestingly, we can use the deterministic approximation in (15)-(18) to characterize the structure of potentially good offer sets. In particular, noting that $p_1 - c_1 > p_2 - c_2 > \ldots > p_n - c_n$, the next proposition shows that the solutions from the deterministic approximation only use sets of the form $\{1, 2, \ldots, i\}$.

Proposition 5 *There exists an optimal solution (w^*, w_0^*) to problem (15)-(18) such that if we define the solution (y^*, x^*) as $x_i^* = \lambda\, w_i^*$ and y^* is as in (9), then the following three statements hold. First, the solution (y^*, x^*) is optimal to the deterministic version of problem (1)-(3) that is obtained by replacing the random variable $\mathsf{Pois}\big(\sum_{S \subseteq \{1,\ldots,n\}} \lambda\, P_i(S)\, y(S)\big)$ in the objective function with its expectation $\sum_{S \subseteq \{1,\ldots,n\}} \lambda\, P_i(S)\, y(S)$. Second, the solution (y^*, x^*) has only one set $S' \subseteq \{1, \ldots, n\}$ such that $y^*(S') \geq 0$ and we have $y^*(S) = 0$ for all other sets $S \subseteq \{1, \ldots, n\}$. Third, the set S' satisfies $S' = \{1, 2, \ldots, j\}$ for some $j = 0, \ldots, n$.*

Proof. The first statement holds by the discussion that follows the formulation of problem (15)-(18). Assuming that (w^*, w_0^*) is an extreme point solution to problem (15)-(18), we focus on the last two statements. By Corollary 8 in Appendix A.2, for the extreme point (w^*, w_0^*) of the set of feasible solutions to problem (15)-(18), there exists a set of products $\tilde{S} \subseteq \{1, \ldots, n\}$ such that $w_i^* = P_i(\tilde{S})$ for all $i = 0, \ldots, n$. By the definition of $P_i(S)$, we have $P_i(\tilde{S})/v_i = P_0(\tilde{S})/v_0$ if $i \in \tilde{S}$ and $P_i(\tilde{S})/v_i = 0$ if $i \notin \tilde{S}$, implying that the extreme point (w^*, w_0^*) satisfies $w_i^*/v_i = w_0^*/v_0$ for all $i \in \tilde{S}$ and $w_i^*/v_i = 0$ for all $i \notin \tilde{S}$. Noting that $p_1 - c_1 > p_2 - c_2 > \ldots > p_n - c_n$, we claim that $\tilde{S}$ has to be a set of the form $\{1, 2, \ldots, j\}$. To see the claim, if the claim does not hold, then we have $k \notin \tilde{S}$ and $i \in \tilde{S}$ for some $k < i$, in which case, $p_k - c_k > p_i - c_i$, $w_k^*/v_k = 0$ and $w_i^*/v_i = w_0^*/v_0$. So, we can increase the value of the decision variable w_k by a small amount and decrease the value of the decision variable w_i by the same small amount to obtain a solution to problem (15)-(18) that provides a strictly better objective value than the solution (w^*, w_0^*). Thus, our claim holds and we get

$$\frac{w_0^*}{v_0} = \frac{w_1^*}{v_1} = \ldots = \frac{w_j^*}{v_j} \geq \frac{w_{j+1}^*}{v_{j+1}} = \frac{w_{j+2}^*}{v_{j+2}} = \ldots = \frac{w_n^*}{v_n} = 0$$

for some $j = 0, \ldots, n$. Using the chain of inequalities above, since $y^*(S)$ is defined as in (9), the solution y^* satisfies $y^*(\{1, 2, \ldots, j\}) \geq 0$ and we have $y^*(S) = 0$ for all other sets of products. $\square$

Proposition 5 motivates working with sets that include a certain number of products with the largest margins. We call such sets as margin ordered sets. we note that there are only $1 + n$ possible margin ordered sets and if we focus only on such sets, then problem (1)-(3) has $1 + 2n$ decision variables. Proposition 4 indicates that such sets can perform well especially when the expected number of customer arrivals is large. Our computational experiments confirm this expectation.

6 Normal Approximation

Although the deterministic approximation we study in the previous section may perform well when the customer volume is large, its performance may not be quite satisfactory when we deal with a small number of customer arrivals. In this section, we develop a different approximate method for problem (4)-(7) that is based on approximating the random variable $\mathsf{Pois}(\lambda\, w_i)$ with a normal random variable with mean $\lambda\, w_i$ and standard deviation $\sqrt{\lambda\, w_i}$.

For brevity, we use $\mathsf{Norm}(\lambda\, w_i)$ to denote the normal random variable with mean $\lambda\, w_i$ and standard deviation $\sqrt{\lambda\, w_i}$. We begin with a few manipulations in the objective function of problem (4)-(7) when we replace $\mathsf{Pois}(\lambda\, w_i)$ with $\mathsf{Norm}(\lambda\, w_i)$. In this problem, the best value of the decision variable x_i for a given w_i is obtained by maximizing $p_i\, \mathbb{E}\{\min\{\mathsf{Norm}(\lambda\, w_i), x_i\}\} - c_i\, x_i$ over x_i. The optimal solution to the last problem is $x_i = \lambda\, w_i + \Phi^{-1}(1 - c_i/p_i)\, \sqrt{\lambda\, w_i}$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal distribution function. Letting $\rho_i = \Phi^{-1}(1 - c_i/p_i)$ for notational brevity and N be the standard normal random variable, replacing the Poisson random variable in the objective function of problem (4)-(7) with the corresponding normal random variable and using the best value of x_i, we have

$$
\begin{aligned}
&p_i\, \mathbb{E}\big\{\min\big\{\mathsf{Norm}(\lambda\, w_i), \lambda\, w_i + \rho_i \sqrt{\lambda\, w_i}\big\}\big\} - c_i\, (\lambda\, w_i + \rho_i \sqrt{\lambda\, w_i}) \\
&\qquad = p_i\, \mathbb{E}\big\{\min\big\{\lambda\, w_i + N\, \sqrt{\lambda\, w_i}, \lambda\, w_i + \rho_i \sqrt{\lambda\, w_i}\big\}\big\} - c_i\, (\lambda\, w_i + \rho_i \sqrt{\lambda\, w_i}) \\
&\qquad\qquad = (p_i - c_i)\, \lambda\, w_i + p_i\, \sqrt{\lambda\, w_i}\, \mathbb{E}\big\{\min\{N, \rho_i\}\big\} - c_i\, \rho_i\, \sqrt{\lambda\, w_i}.
\end{aligned}
$$

For the third term on the right side of the chain of equalities above, using $\Phi(\cdot)$ and $\phi(\cdot)$ to respectively denote the standard normal distribution and density functions, a well-known computation shows that $\mathbb{E}\{\min\{N, \rho_i\}\} = \rho_i\, (1 - \Phi(\rho_i)) - \phi(\rho_i) = \rho_i\, (1 - \Phi(\Phi^{-1}(1 - c_i/p_i))) - \phi(\rho_i) = \rho_i\, c_i/p_i - \phi(\rho_i)$; see Johnson and Montgomery (1974). Thus, the expression on the right side of the chain of equalities above is $(p_i - c_i)\, \lambda\, w_i + p_i\, \sqrt{\lambda\, w_i}\, (\rho_i\, c_i/p_i - \phi(\rho_i)) - c_i\, \rho_i\, \sqrt{\lambda\, w_i} = (p_i - c_i)\, \lambda\, w_i - p_i\, \phi(\rho_i)\, \sqrt{\lambda\, w_i}$. So, when we approximate the Poisson random variable with the corresponding normal random variable, problem (4)-(7) is equivalent to

$$
\begin{aligned}
\max \quad & \sum_{i=1}^{n} \Big\{ (p_i - c_i)\, \lambda\, w_i - p_i\, \phi(\rho_i)\, \sqrt{\lambda\, w_i} \Big\} && (20) \\
\text{subject to} \quad & (5), (6), (w, w_0) \geq 0. && (21)
\end{aligned}
$$

Similar to problem (4)-(7), the objective function of the problem above decomposes by the products and we can obtain an optimal solution by using a dynamic program. As we show in our computational experiments, the solution obtained from the problem above can perform remarkably well even when the deterministic approximation given in the previous section turns out to be unsatisfactory.

One of the welcome features of the normal approximation is that its optimal solution offers a single set over the whole selling horizon. This is to say that although our model allows offering different sets over the selling horizon and the durations of time over which the different sets are offered are decision variables, the optimal solution to the normal approximation chooses one set and offers this set throughout whole selling horizon. We focus on this result in the next proposition.

Proposition 6 *There exists an optimal solution* (w^*, w_0^*) *to problem* (20)-(21) *such that if we define the solution* (y^*, x^*) *as* $x_i^* = \lambda\, w_i^* + \rho_i \sqrt{\lambda\, w_i^*}$ *and* y^* *is as in* (9)*, then the following two statements hold. First, the solution* (y^*, x^*) *is optimal to the version of problem* (1)-(3) *that is obtained by replacing the Poisson random variable* $\mathsf{Pois}\big(\sum_{S \subseteq \{1,\ldots,n\}} \lambda\, P_i(S)\, y(S)\big)$ *in the objective function with the normal random variable* $\mathsf{Norm}\big(\sum_{S \subseteq \{1,\ldots,n\}} \lambda\, P_i(S)\, y(S)\big)$*. Second, the solution* (y^*, x^*) *has only one set* $S' \subseteq \{1, \ldots, n\}$ *such that* $y^*(S') \geq 0$ *and we have* $y^*(S) = 0$ *for all other sets* $S \subseteq \{1, \ldots, n\}$*.*

Proof. The first statement can be shown by using the same argument in the proof of Theorem 1. We focus on the second statement. By Corollary 8 in Appendix A.2, for any extreme point $(\tilde{w}, \tilde{w}_0)$ of the set of feasible solutions to problem (20)-(21), there exists a set of products $\tilde{S} \subseteq \{1, \ldots, n\}$ such that $\tilde{w}_i = P_i(\tilde{S})$ for all $i = 0, \ldots, n$. As mentioned in the proof of Proposition 5, we have $P_i(\tilde{S})/v_i = P_0(\tilde{S})/v_0$ if $i \in \tilde{S}$ and $P_i(\tilde{S})/v_i = 0$ if $i \notin \tilde{S}$, implying that any extreme point $(\tilde{w}, \tilde{w}_0)$ of the set of feasible solutions to problem (20)-(21) satisfies $\tilde{w}_i/v_i \in \{0, \tilde{w}_0/v_0\}$ for all $i = 1, \ldots, n$. Since problem (20)-(21) is maximizing a convex function, its optimal objective value is achieved at an extreme point. Therefore, we can assume that the optimal solution (w^*, w_0^*) to problem (20)-(21) satisfies $w_i^*/v_i \in \{0, w_0^*/v_0\}$ for all $i = 1, \ldots, n$. Without loss of generality, reindexing the products, for some $j = 0, \ldots, n$, we have $w_i^*/v_i = w_0^*/v_0$ for all $i = 1, \ldots, j$ and $w_i^*/v_i = 0$ for all $i = j+1, \ldots, n$. We write the last fact as

$$\frac{w_0^*}{v_0} = \frac{w_1^*}{v_1} = \ldots = \frac{w_j^*}{v_j} \geq \frac{w_{j+1}^*}{v_{j+1}} = \ldots = \frac{w_n^*}{v_n} = 0.$$

Since $y^*(S)$ is defined as in (9), noting the chain of inequalities above, the solution y^* satisfies $y^*(\{1, 2, \ldots, j\}) \geq 0$ and $y^*(S) = 0$ for all other sets $S \subseteq \{1, \ldots, n\}$. $\square$

As mentioned above, the normal approximation has remarkably good numerical performance. Given that the normal approximation also offers a single set, it can be practically attractive.

7 Integer Programming Approximation

Due to the nonlinear and convex objective function of problem (20)-(21), solving this problem directly can be difficult and it is essential to use a dynamic program instead, which comes with the necessity of discretizing the state and action variables and creating custom code. In this section, we develop a simple integer program that can obtain quite good solutions to problem (20)-(21). By solving this integer problem through commercial integer programming packages, we can altogether avoid the necessity to use a dynamic program to solve problem (20)-(21). To formulate the integer program, we let $g_i(w_i) = (p_i - c_i)\,\lambda\, w_i - p_i\,\phi(\rho_i)\sqrt{\lambda\, w_i}$ so that the objective function of problem (20)-(21) is $\sum_{i=1}^n g_i(w_i)$. The function $g_i(\cdot)$ is convex and it satisfies $g(w_i) = 0$ at $w_i = 0$ and at $w_i = K_i$, where K_i is defined as $K_i = \frac{1}{\lambda}\big(\frac{p_i\,\phi(\rho_i)}{p_i - c_i}\big)^2$. For $w_i \in [0, K_i]$, we have $g_i(w_i) \leq 0$. This behavior of $g_i(\cdot)$ is shown in Figure 1.a, where we plot this function for different choices of λ, p_i and c_i.

We make two observations. The first observation is that an optimal solution (w^*, w_0^*) to problem (20)-(21) would never have $g_i(w_i^*) < 0$ for some $i = 1, \ldots, n$. To see this claim, if we have $g_i(w_i^*) < 0$, then we can decrease the value of w_i^* to zero and increase the value of w_0^* by w_i^* to obtain a new solution. The new solution is still feasible to problem (20)-(21). Furthermore, since $g_i(0) = 0$ but $g_i(w_i^*) < 0$, the new solution provides a strictly better objective value to problem (20)-(21) than does the solution (w^*, w_0^*), establishing the claim. Note that problem (20)-(21) maximizes $\sum_{i=1}^n g_i(w_i)$ over $(w, w_0) \geq 0$ subject to constraints (5) and (6). Thus, since an optimal solution (w^*, w_0^*) to problem (20)-(21) would never have $g_i(w_i^*) < 0$, problem (20)-(21) is equivalent to maximizing $\sum_{i=1}^n [g_i(w_i)]^+$ over $(w, w_0) \geq 0$ subject to constraints (5) and (6). We use $[a]^+$ to denote $\max\{a, 0\}$.

The second observation is that the derivative of $g_i(\cdot)$ at K_i is $\frac{1}{2}(p_i - c_i)\,\lambda$. So, since $g_i(\cdot)$ is convex and $g_i(K_i) = 0$, the hockey stick function $\frac{1}{2}(p_i - c_i)\,\lambda\,[w_i - K_i]^+$ lower bounds $[g_i(w_i)]^+$. In Figure 1.b, we plot $[g_i(\cdot)]^+$ and $\frac{1}{2}(p_i - c_i)\,\lambda\,[\cdot - K_i]^+$ for a particular choice of λ, p_i and c_i. The hockey stick function $\frac{1}{2}(p_i - c_i)\,\lambda\,[\cdot - K_i]^+$ is a good approximation to $[g_i(\cdot)]^+$ especially when the behavior of $g_i(\cdot)$ to the right of K_i is linear. In this case, we propose solving the problem

$$\max \quad \sum_{i=1}^n \frac{1}{2}(p_i - c_i)\,\lambda\,[w_i - K_i]^+ \tag{22}$$

$$\text{subject to} \quad (5), (6), (w, w_0) \geq 0 \tag{23}$$

as an approximation to problem (20)-(21). The problem above maximizes the sum of n piecewise linear convex functions. So, it can be formulated as an integer program. Since there is only one point of nondifferentiability in the piecewise linear convex functions, this requires using n additional binary decision variables. Although integer programming is NP-hard in general, our experience is that commercial solvers, such as Gurobi 4.5, can solve instances of problem (22)-(23) quite fast.

In our computational experiments, we demonstrate that problem (22)-(23) can obtain quite accurate solutions to problem (20)-(21) over a range of parameter values. An important observation is that if we replace $[w_i - K_i]^+$ in the objective function of problem (22)-(23) with $[w_i - K_i]$, then problem (22)-(23) has the same optimal solution as the deterministic approximation in (15)-(18). However, the solutions obtained from problem (22)-(23) generally perform significantly better than those obtained from problem (15)-(18). Thus, it is interesting that the presence of the $[\cdot]^+$ operator makes such a drastic change in the quality of the solutions obtained from problem (22)-(23).

As mentioned a moment ago, our computational work indicates that problem (22)-(23) can obtain quite accurate solutions to problem (20)-(21). A natural question is whether we can make this empirical observation more precise. In the next proposition, we indeed show that if we solve problem (22)-(23) and use this solution as a possible solution to problem (20)-(21), then we do not suffer a loss in the expected profit by more than 50%. We emphasize that while this proposition bounds the loss by 50%, the loss in the expected profit turns out to be far less than 50% in our computational work.

Proposition 7 *Recalling that* $\sum_{i=1}^n g_i(\cdot)$ *is the objective function of problem* (20)-(21)*, if we let* (w^*, w_0^*) *be an optimal solution to problem* (20)-(21) *and* $(\hat{w}, \hat{w}_0)$ *be an optimal solution to problem* (22)-(23)*, then we have* $2\sum_{i=1}^n g_i(\hat{w}_i) \geq \sum_{i=1}^n g_i(w_i^*) \geq \sum_{i=1}^n g_i(\hat{w}_i)$*.*

Proof. The second inequality in the proposition follows from the fact that $(\hat{w}, \hat{w}_0)$ is a feasible but not necessarily an optimal solution to problem (20)-(21). We focus on the first inequality. The sets of feasible solutions to problems (20)-(21) and (22)-(23) are the same and we use $\mathcal{W}$ to denote this common set. In this case, problem (20)-(21) is equivalent to $\max_{(w,w_0)\in\mathcal{W}} \sum_{i=1}^n g_i(w_i) = \max_{(w,w_0)\in\mathcal{W}} \sum_{i=1}^n [g_i(w_i)]^+$, where the equality follows from the first observation above right before formulating problem (22)-(23). On the other hand, the derivative of $g_i(\cdot)$ is given by $(p_i - c_i)\,\lambda - \frac{p_i\,\phi(\rho_i)\,\sqrt{\lambda}}{2\sqrt{w_i}}$, which is always smaller than $(p_i - c_i)\,\lambda$. This implies that the hockey stick function $(p_i - c_i)\,\lambda\,[w_i - K_i]^+$ upper bounds $[g_i(w_i)]^+$. Thus, we obtain

$$\begin{aligned}\sum_{i=1}^n g_i(w_i^*) &= \max_{(w,w_0)\in\mathcal{W}} \left\{ \sum_{i=1}^n g_i(w_i) \right\} = \max_{(w,w_0)\in\mathcal{W}} \left\{ \sum_{i=1}^n [g_i(w_i)]^+ \right\} \\ &\leq \max_{(w,w_0)\in\mathcal{W}} \left\{ \sum_{i=1}^n (p_i - c_i)\,\lambda\,[w_i - K_i]^+ \right\} = 2 \max_{(w,w_0)\in\mathcal{W}} \left\{ \sum_{i=1}^n \frac{1}{2}\,(p_i - c_i)\,\lambda\,[w_i - K_i]^+ \right\} \\ &= 2\sum_{i=1}^n \frac{1}{2}\,(p_i - c_i)\,\lambda\,[\hat{w}_i - K_i]^+ \leq 2\sum_{i=1}^n g_i(\hat{w}_i),\end{aligned}$$

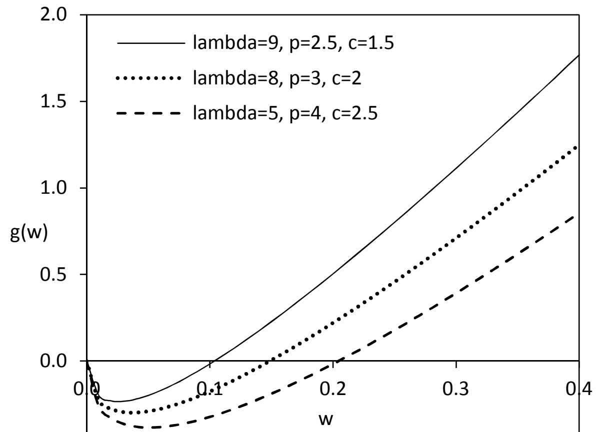
where the last equality is by the fact that $(\hat{w}, \hat{w}_0)$ is an optimal solution to problem (22)-(23). The last inequality follows by the fact $\frac{1}{2}\,(p_i - c_i)\,\lambda\,[w_i - K_i]^+$ lower bounds $[g_i(w_i)]^+$ and if $g_i(\hat{w}_i) < 0$, then it must be the case that $\hat{w}_i < K_i$ so that we can decrease the value of $\hat{w}_i$ to zero without changing the objective value provided by the solution $(\hat{w}, \hat{w}_0)$ for problem (22)-(23). $\square$

We note that the analysis in the proof above is not completely tight. In particular, the derivative of $g_i(\cdot)$ is given by $(p_i - c_i)\,\lambda - p_i\,\phi(\rho_i)\,\sqrt{\lambda}/(2\sqrt{w_i})$. Since $w_i \in [0,1]$, this derivative can be upper bounded by $(p_i - c_i)\,\lambda - p_i\,\phi(\rho_i)\,\sqrt{\lambda}/2$. By using the latter upper bound on the derivative of $g_i(\cdot)$ instead of the upper bound $(p_i - c_i)\,\lambda$ used in the proof, we can repeat the analysis. In the end, we get a tighter bound than two, but this bound depends on the problem parameters and it is not as clean as two.

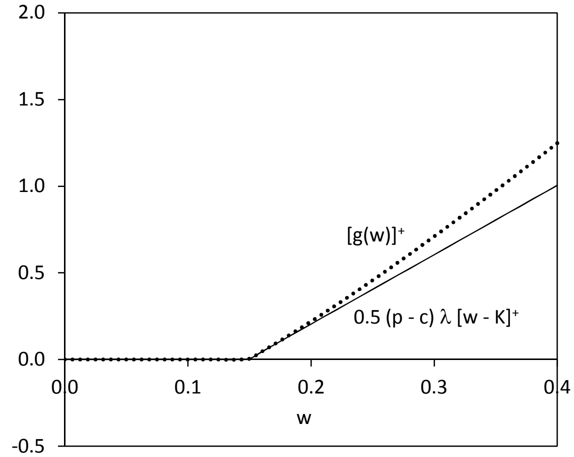
Finally, since problem (22)-(23) maximizes a convex objective function, we can repeat the proof of Proposition 6 to show that an optimal solution from problem (22)-(23) offers a single set over the whole selling horizon, similar to the optimal solution from the normal approximation.

8 Computational Experiments

In this section, we test the performance of the stocking and product offer decisions that we obtain by solving problem (1)-(3) and by solving various approximations to this problem.



(a)



(b)

Figure 1: The function $g_i(\cdot)$ in (a) and the functions $[g_i(\cdot)]^+$ and $\frac{1}{2}(p_i - c_i)\,\lambda\,[\cdot - K_i]^+$ in (b).

8.1 Benchmark Strategies

If the product offer decisions that we make over the selling horizon are characterized by the decision variables $y = \{y(S) : S \subseteq \{1, \ldots, n\}\}$, then we can choose the stocking quantities by solving $\max_{x_i \geq 0} p_i \, \mathbb{E}\{\min\{\mathsf{Pois}\big(\sum_{S \subseteq \{1,\ldots,n\}} \lambda \, P_i(S) \, y(S)\big), x_i\}\} - c_i \, x_i$ for all $i = 1, \ldots, n$. All of our benchmark strategies choose the stocking quantities by solving the last problem, but they differ in the way they choose the values of the decision variables y. We test the performance of five benchmark strategies.

Stochastic Model (STO). This benchmark strategy corresponds to the model that we study in this paper. STO solves problem (1)-(3) by using the dynamic programming formulation in (10)-(13) and uses the optimal values of the decision variables y to decide which sets of products to offer over the selling horizon. Noting that the state variable (z_i, w_0) in the dynamic program takes values over $[0, 1]^2$ and the action variable w_i takes values over $[0, 1]$, we divide the interval $[0, 1]$ into 1,000 subintervals and solve the dynamic program approximately by using the discrete state and action variables.

Deterministic Approximation (DET). This benchmark strategy solves the deterministic approximation given in problem (15)-(18) to obtain the optimal values of the decision variables (w, w_0), which we denote by (w^*, w_0^*). In this case, DET uses the transformation in (9) to figure out the duration of time over which each set of products should be offered.

Margin Ordered Sets (MOR). This benchmark strategy is motivated by Proposition 5, which shows that the solution to the deterministic approximation uses only margin ordered sets. Thus, the idea behind MOR is to offer the best margin ordered set. In particular, MOR orders the products such that we have $p_1 - c_1 \geq p_2 - c_2 \geq \ldots \geq p_n - c_n$ and considers offering the set of products $S_i = \{1, 2, \ldots, i\}$ over the whole selling horizon by setting $y(S_i) = 1$ and $y(S) = 0$ for all $S \neq S_i$. For all $i = 1, \ldots, n$, MOR checks the expected profit obtained by offering only the set S_i and chooses the set that provides the largest expected profit.

Normal Approximation (NAX). This benchmark strategy builds on the normal approximation given in problem (20)-(21). Similar to DET, once NAX obtains the optimal values of the decision variables (w, w_0), denoted by (w^*, w_0^*), it uses the transformation in (9) to figure out the duration of time over which each set of products should be offered.

Integer Programming Approximation (INP). This benchmark strategy uses problem (22)-(23) to obtain the optimal values of the decision variables (w, w_0). As mentioned in Section 7, since problem (22)-(23) maximizes the sum of piecewise linear functions, this problem can be formulated as an integer program. Similar to DET and NAX, using (w^*, w_0^*) to denote an optimal solution to problem (22)-(23), INP uses the transformation in (9) to compute the duration of time over which each set of products should be offered.

8.2 Experimental Setup

In our experimental setup, we enrich the cost structure of our model to generate test problems with a variety of attributes. The cost components in problem (1)-(3) correspond to the expected revenue from the sales and the stocking cost. In our computational experiments, we assume that if a customer chooses a product for which we do not have any stock, then we incur an additional cost, which can be visualized as the cost associated with making an emergency procurement or the cost associated with the loss of goodwill. Letting $\bar{c}_i$ be the such cost of making an emergency procurement for product i, we incorporate the new cost component by subtracting the expression

$$\sum_{i=1}^{n} \bar{c}_i \, \mathbb{E}\Big\{ \max \Big\{ \mathsf{Pois}\big(\textstyle\sum_{S \subseteq \{1,\ldots,n\}} \lambda \, P_i(S) \, y(S)\big) - x_i, 0 \Big\}\Big\}$$

from the objective function of problem (1)-(3). Noting that $\max\{\mathsf{Pois}(\sum_{S \subseteq \{1,\ldots,n\}} \lambda \, P_i(S) \, y(S)) - x_i, 0\} = \mathsf{Pois}(\sum_{S \subseteq \{1,\ldots,n\}} \lambda \, P_i(S) \, y(S)) - \min\{\mathsf{Pois}(\sum_{S \subseteq \{1,\ldots,n\}} \lambda \, P_i(S) \, y(S)), x_i\}$, subtracting the cost component above from the objective function of problem (1)-(3) does not complicate our model and all of our results continue to hold. In particular, we can obtain a solution to our model by using the transformation in Theorem 1 and the deterministic approximation still uses margin ordered sets. It is also possible to incorporate another cost component for the salvage values of the products that are unsold at the end of the selling horizon and this cost component can be handled in a similar fashion.

In our test problems, we have $n = 20$ products, among which we need to choose sets to offer to our customers. Letting U_i be a sample from the uniform distribution, we set the preference weight of product i as $v_i = 1 + 9\,U_i$ so that the preference weights are uniformly distributed over the interval $[1, 10]$. On the other hand, we set the unit revenue of product i as $p_i = 100 + 400\,(1 - U_i)^2$. Due to the presence of U_i in both the preference weight and unit revenue, the preference weight of product i is negatively correlated with its unit revenue, indicating that more expensive products tend to be less attractive to customers. This setup allows to generate more interesting test problems in the sense that expensive products are not trivially the most desirable ones to include in the offer set and one should carefully weigh the attractiveness of a product against its unit revenue to figure out whether a product should be offered. Squaring the term $1 - U_i$ in the unit revenue of product i ensures that the unit revenues of the products have a skewed distribution. In other words, due to the term $(1 - U_i)^2$, we have few products with large unit revenues and many products with small unit revenues. To come up with the unit costs of the products, we sample χ_i from the uniform distribution over the interval $[0.3, 0.7]$ and set the unit cost of product i as $c_i = \chi_i p_i$. To come up with the emergency procurement cost of product i, we sample ζ_i from the uniform distribution over the interval $[\bar{\zeta} - 0.5, \bar{\zeta} + 0.5]$ and set $\bar{c}_i = \zeta_i p_i$. Thus, larger values of $\bar{\zeta}$ imply larger emergency procurement costs. We vary $\bar{\zeta}$ in our computational experiments. Finally, we choose the preference weight v_0 associated with the no purchase option such that $P_0 = v_0/(v_0 + \sum_{i=1}^n v_i)$. In other words, even if we offer all of the products, then the no choice probability is equal to P_0. Similar to $\bar{\zeta}$, we vary P_0 in our computational experiments.

Recalling that λ is the arrival rate of customers, we vary λ over $\{50, 100, 250\}$, P_0 over $\{0.05, 0.1\}$ and $\bar{\zeta}$ over $\{1, 2, 3\}$. We use the triplet $(\lambda, P_0, \bar{\zeta})$ to denote each combination of these parameters. For each parameter combination, we generate 50 individual problem instances and test the performance of the five benchmark strategies described above on each problem instance.

8.3 Computational Results

Table 1 compares the expected profits corresponding to the five benchmark strategies for each parameter combination. The first column in this table shows the parameter combination in consideration. The second column shows the average expected profit obtained by STO, where average is taken over all 50 problem instances that we generate for a particular parameter combination. In other words, using $\mathsf{Prof}^k_{\mathrm{STO}}$ to denote the expected profit obtained by STO for problem instance k, the second column gives $\frac{1}{50}\sum_{k=1}^{50} \mathsf{Prof}^k_{\mathrm{STO}}$. The third, fourth, fifth and sixth columns follow a similar approach to show the average expected profits corresponding to DET, MOR, NAX and INP. The seventh column in Table 1 shows the average percent gap between the expected profits obtained by STO and DET, where average is, again, taken over all 50 problem instances in a particular parameter combination. So, letting $\mathsf{Prof}^k_{\mathrm{DET}}$ be the expected profit obtained by DET for problem instance k, the seventh column gives

$$\frac{1}{50}\sum_{k=1}^{50} 100\, \frac{\mathsf{Prof}^k_{\mathrm{STO}} - \mathsf{Prof}^k_{\mathrm{DET}}}{\mathsf{Prof}^k_{\mathrm{STO}}}.$$

In a similar fashion, the eighth, ninth and tenth columns compare the expected profits obtained by STO and the remaining three benchmark strategies, MOR, NAX and INP. The eleventh column in Table 1 shows the 90th percentile of the performance gaps between STO and DET. In particular, this column gives the 90th percentile of the data $\{100\,(\mathsf{Prof}^k_{\mathrm{STO}} - \mathsf{Prof}^k_{\mathrm{DET}})/\mathsf{Prof}^k_{\mathrm{STO}} : k = 1, \ldots, 50\}$. The goal of this column is to give an indication of how large the performance gaps between STO and DET can get in worst problem instances. Similarly, the twelfth, thirteenth and fourteenth columns give the 90th percentiles of the performance gaps between STO and the remaining benchmark strategies.

The results in Table 1 indicate that STO can provide significant improvements over DET, especially when the arrival rate of customers is small or the emergency procurement cost is large. The coefficient of variation of a Poisson random variable with mean λ is $1/\sqrt{\lambda}$. Therefore, the parameter combinations with small arrival rates correspond to cases where the coefficient of variation for customer arrivals is large. DET ignores the uncertainty in customer arrivals and it is not surprising that the performance gap between STO and DET gets larger as λ decreases and the coefficient of variation gets larger. Considering the effect of emergency procurement cost, as this cost decreases, we are more likely to use emergency procurements, rather than purchasing at the beginning of the selling horizon. Indeed, if the emergency procurement cost is smaller than the purchasing cost, then it is optimal to satisfy the demand only through emergency procurements and the stocking decisions become trivial. Thus, it is understandable that using a sophisticated method such as STO provides significant improvements over DET when emergency procurements are more costly. Also, for all of the parameter combinations, the average performance gap between STO and DET gets larger as P_0 increases. We note that the preference weight associated with the no purchase option gets larger as P_0 increases. If the preference weight associated with the no purchase option is larger, then customers have larger tendency to leave without purchasing anything and it is desirable to offer a larger set of products to counteract this tendency. When we offer a larger set of products, the total stock has to be split among a larger number of products and the stocking level for each product decreases. When dealing with small stocking levels, stocking and product offer decisions become more delicate and STO tends to provide larger improvements over DET. Overall, for the parameter combinations with large arrival rates, small emergency procurement costs and small no purchase preference weights, the average performance gap between STO and DET is smaller than 3%, but the average performance gap between the two benchmark strategies can still exceed 60% on the other side of the spectrum. The performance of MOR is significantly better than that of DET. Despite the fact that both DET and MOR work with margin ordered sets, the deterministic approximation used by DET is not able to pinpoint the best margin ordered set. By making an exhaustive check over all possible margin ordered sets, MOR can perform significantly better than DET. Still, MOR can suffer from the same shortcomings as DET in the sense that the performance of MOR can deteriorate when the arrival rate of customers gets small or the emergency procurement cost gets large. For the parameter combinations with small arrival rates and large emergency procurement costs, MOR may lag behind STO by more than 20%.

The performance is NAX is substantially more satisfactory than that of DET or MOR. In all of the parameter combinations, the average performance gap between NAX and STO is less than a percent. Similar to DET and MOR, the performance gap between NAX and STO increases as the arrival rate of customers gets smaller, or the emergency procurement cost gets larger, or the preference weight of the no purchase option gets larger, but NAX is far less sensitive to these problem parameters when compared with DET or MOR. Even for the worst parameter combination, the performance gap between NAX and STO is less than half a percent. Finally, we observe that NIP can perform quite well, yielding average performance gaps with STO that are less than 2%. Even for the most problematic parameter combination, when DET and MOR give average performance gaps respectively on the order of 60% and 20%, the average performance gap of INP is 1.65%.

A tempting approach for solving problem (4)-(7) involves assuming that the stocking quantities x are fixed and finding the best values of the decision variables (w, w_0) for the fixed stocking quantities. The objective function of problem (4)-(7) is concave in (w, w_0) so that we can solve this problem for the fixed stocking quantities by using standard convex optimization tools. In this case, we can try to use marginal analysis to choose products whose stocking quantities should be increased or decreased to improve the

Params. $(\lambda,\ P_0,\ \bar{\zeta})$	Expected Profit STO	DET	MOR	NAX	INP	Avg. Gap with STO DET	MOR	NAX	INP	90th Perc. Gap with STO DET	MOR	NAX	INP
$(50, 0.05, 1)$	4,424	3,623	4,192	4,420	4,370	18.54	5.51	0.10	1.21	33.71	13.56	0.33	3.27
$(50, 0.05, 2)$	4,127	2,963	3,804	4,124	4,069	28.99	8.24	0.09	1.35	49.81	20.53	0.35	4.00
$(50, 0.05, 3)$	3,948	2,529	3,557	3,942	3,869	36.99	10.39	0.17	1.93	67.05	25.85	0.62	4.86
$(50, 0.10, 1)$	3,061	2,212	2,751	3,058	3,035	28.52	10.58	0.07	0.88	47.24	22.06	0.24	2.99
$(50, 0.10, 2)$	2,773	1,497	2,345	2,769	2,726	47.36	16.16	0.13	1.68	77.64	34.63	0.28	4.88
$(50, 0.10, 3)$	2,606	1,023	2,098	2,598	2,562	62.67	20.36	0.31	1.65	*100.17	43.00	1.10	3.50
$(100, 0.05, 1)$	10,066	9,244	9,833	10,064	9,974	8.31	2.42	0.02	0.92	15.92	7.16	0.06	2.51
$(100, 0.05, 2)$	9,568	8,352	9,236	9,564	9,457	13.00	3.64	0.04	1.14	22.85	10.81	0.12	3.14
$(100, 0.05, 3)$	9,262	7,766	8,862	9,256	9,137	16.59	4.57	0.06	1.35	29.07	11.57	0.25	3.07
$(100, 0.10, 1)$	7,298	6,459	6,980	7,295	7,247	11.76	4.53	0.04	0.71	23.29	11.59	0.12	2.16
$(100, 0.10, 2)$	6,806	5,501	6,331	6,801	6,736	19.69	7.33	0.08	1.06	35.58	16.05	0.39	2.52
$(100, 0.10, 3)$	6,503	4,871	5,936	6,497	6,425	25.83	9.17	0.11	1.24	45.73	20.09	0.41	3.20
$(250, 0.05, 1)$	28,287	27,473	28,097	28,287	28,100	2.90	0.68	0.00	0.66	5.04	2.36	0.01	1.82
$(250, 0.05, 2)$	27,359	26,123	27,052	27,358	27,143	4.59	1.16	0.00	0.81	7.81	3.77	0.01	2.45
$(250, 0.05, 3)$	26,795	25,242	26,395	26,793	26,549	5.91	1.56	0.01	0.96	10.10	4.91	0.02	2.53
$(250, 0.10, 1)$	21,344	20,615	21,146	21,341	21,180	3.44	0.95	0.02	0.80	8.48	2.94	0.07	2.15
$(250, 0.10, 2)$	20,389	19,167	19,988	20,384	20,250	6.09	2.06	0.03	0.74	13.46	5.22	0.13	2.49
$(250, 0.10, 3)$	19,815	18,228	19,258	19,813	19,678	8.16	2.95	0.01	0.70	16.74	7.32	0.02	1.94

Table 1: Expected profits obtained by STO, DET, MOR, NAX and INP. (*A performance gap more than 100% indicates that the benchmark strategy obtains a negative expected profit.)

objective value of problem (4)-(7). Due to nontrivial interactions between the decision variables x and (w, w_0), this approach does not yield satisfactory results. To give an example, we consider the case with a single product, $\lambda = 100$, $p_1 = 130$, $c_1 = 60$, $\bar{c}_1 = 220$, $v_0 = 32$ and $v_1 = 8$. Given that we fix the stocking quantity of the product at x_1, we let $F(x_1)$ be the optimal objective value of problem (4)-(7). Figure 2 plots $F(x_1)$ as a function of x_1, indicating that $F(x_1)$ immediately decreases as x_1 increases starting from zero. Thus, if we start with a zero stocking quantity and use marginal analysis to find the best stocking quantity, then we get stuck at zero. This is not an uncommon occurrence. For many of our test problems, we get stuck at zero stocking quantities by using marginal analysis starting from zero. We tried initializing our stocking quantities at larger values, but even in this case, we got stuck at stocking quantities that are less satisfactory than the ones provided by NAX or INP.

Table 2 provides summary statistics for the decisions made by the five benchmark strategies. The first column in this table shows the parameter combination in consideration. The second to sixth columns show the total stocking quantities for the benchmark strategies. Letting x be the stocking decisions made by a benchmark strategy, the total stocking quantity is simply $\sum_{i=1}^{n} x_i$. The seventh to eleventh columns show the total expected demand induced by the product offer decisions of STO, DET, MOR, NAX and INP. Using y to denote the product offer decisions made by a benchmark strategy, the total expected demand is given by $\lambda \sum_{S \subseteq \{1,\dots,n\}} [1 - P_0(S)]\, y(S)$, where the summation can be interpreted as the probability that a customer chooses a product. The twelfth to sixteenth columns show the average number of products offered, which is computed as $\sum_{S \subseteq \{1,\dots,n\}} |S|\, y(S)$. Similar to Table 1, Table 2 gives the results averaged over all 50 problem instances in a particular parameter combination.

From Table 2, we observe that as P_0 increases, the no purchase preference weight gets larger and customers are more likely to leave without making a choice. The benchmark strategies counteract

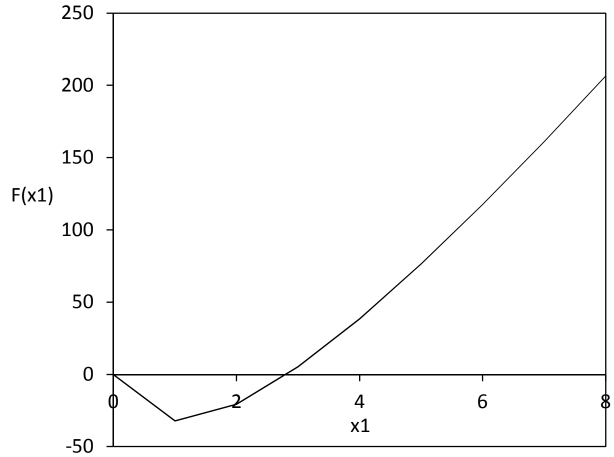


Figure 2: Optimal objective value of problem (4)-(7) for given stocking quantities.

this tendency by increasing the average number of offered products so that customers are more likely to make a purchase within a larger set. Despite the fact that we offer a larger set of products, the expected demand induced by the benchmark strategies still decreases when the preference weight of the no purchase option gets larger. Following the decrease in the expected demand, the stocking quantities also get smaller. As the cost of emergency procurement increases, it is more preferable to purchase the products at the beginning of the selling horizon and the total stocking quantities increase. Comparing the benchmark strategies among each other, DET stands out as an outlier as it tries to induce noticeably larger total expected demand and uses larger stocking quantities. The average number of products offered by DET is also significantly larger than the other benchmark strategies. Finally, the total expected demand induced and the average number of products offered by DET do not depend on the emergency procurement cost. Since DET uses only the expected values of the demand, it does not plan for emergency procurements at all by trying to serve the total expected demand exactly.

Overall, our results indicate that the solution strategies inspired by deterministic approximations can work reasonably well when we have a large volume of customers arriving into the system and the cost of correcting an error in the stocking decisions is not prohibitive. This is evidenced by the respectable performance of DET and MOR with large values of λ and small values of $\bar{\zeta}$. However, as the demand variability and the emergency procurement cost increase, we see significant gaps between a deterministic and a stochastic model. As demonstrated by NAX, over a range of parameter combinations, approximating the number of customer arrivals by using a normal distribution works quite well. In addition, the solutions provided by NAX can be particularly attractive since they offer a single set over the whole selling horizon. Noting that solving the model with the normal approximation still requires working with a dynamic program, it turns out that we can formulate a simple integer program, as in INP, that can closely track the solutions to the dynamic program. We can solve this integer program directly by using commercial integer programming packages. Similar to the normal approximation, the solutions from the integer program also offer a single set.

Params.	Stocking Quantity					Expected Demand					Average Number of Offered Products				
$(\lambda,\ P_0,\ \bar{\zeta})$	STO	DET	MOR	NAX	INP	STO	DET	MOR	NAX	INP	STO	DET	MOR	NAX	INP
$(50, 0.05, 1)$	41.04	48.56	40.38	41.52	43.60	32.31	37.42	31.47	32.51	33.92	3.39	6.56	4.04	3.55	4.16
$(50, 0.05, 2)$	43.26	53.50	41.46	43.56	46.18	31.63	37.42	30.00	31.82	33.40	3.07	6.56	3.66	3.23	3.78
$(50, 0.05, 3)$	43.52	56.52	41.66	44.76	47.98	30.91	37.42	28.80	31.52	33.33	2.83	6.56	3.38	3.07	3.68
$(50, 0.10, 1)$	37.68	45.50	34.90	38.16	39.58	28.30	33.96	25.97	28.60	29.59	4.39	8.60	4.98	4.53	5.00
$(50, 0.10, 2)$	39.42	51.02	35.30	39.84	42.34	27.43	33.96	24.05	27.61	29.10	3.87	8.60	4.44	4.03	4.72
$(50, 0.10, 3)$	40.10	54.46	35.96	41.36	43.48	26.86	33.96	23.17	27.50	28.61	3.53	8.60	4.20	3.83	4.40
$(100, 0.05, 1)$	80.54	91.16	80.00	80.96	85.64	66.94	74.83	66.23	67.19	70.65	4.07	6.56	4.58	4.17	4.92
$(100, 0.05, 2)$	84.30	97.74	81.72	85.32	88.54	66.33	74.83	63.83	67.03	69.03	3.81	6.56	4.18	3.93	4.56
$(100, 0.05, 3)$	86.18	101.72	83.92	87.94	91.00	65.72	74.83	63.39	66.79	68.60	3.63	6.56	4.10	3.83	4.42
$(100, 0.10, 1)$	74.20	84.56	71.68	74.94	78.16	59.83	67.93	57.62	60.26	62.69	5.22	8.60	6.04	5.40	6.16
$(100, 0.10, 2)$	77.02	92.06	73.24	77.74	81.98	58.16	67.93	54.71	58.61	61.45	4.79	8.60	5.48	4.94	5.68
$(100, 0.10, 3)$	78.70	97.00	74.32	80.06	84.36	57.10	67.93	53.22	58.01	60.73	4.56	8.60	5.20	4.75	5.42
$(250, 0.05, 1)$	198.44	212.68	197.00	198.40	206.48	174.94	187.09	173.73	174.93	181.84	4.84	6.56	5.18	4.84	5.70
$(250, 0.05, 1)$	202.70	223.24	201.48	204.26	213.52	171.81	187.09	170.37	173.02	180.12	4.55	6.56	4.92	4.63	5.44
$(250, 0.05, 1)$	204.20	229.40	204.96	205.78	217.54	169.20	187.09	169.23	170.37	179.04	4.33	6.56	4.82	4.41	5.28
$(250, 0.10, 1)$	180.22	196.98	179.96	180.92	188.96	155.56	169.81	155.21	156.07	163.22	6.25	8.60	6.96	6.39	7.46
$(250, 0.10, 2)$	186.92	208.22	185.92	189.16	194.90	153.68	169.81	152.40	155.39	159.91	5.91	8.60	6.72	6.11	6.88
$(250, 0.10, 3)$	190.02	215.58	188.56	190.90	199.58	152.08	169.81	149.99	152.67	158.97	5.72	8.60	6.50	5.77	6.60

Table 2: Summary statistics for the decisions made by STO, DET, MOR, NAX and INP.

9 Conclusion and Discussion

We studied a joint stocking and product offer problem under the multinomial logit model. Our model builds on the one in van Ryzin and Mahajan (1999), but by allowing different sets of products being offered over the selling horizon, we avoid the combinatorial aspects of the product offer problem and accommodate products with arbitrary unit revenues and costs. Although a naive formulation of our model results in a difficult nonlinear program, we transformed this nonlinear program into an equivalent one that is more tractable. We developed a solution approach that exploits the separability of the objective function by using dynamic programming. Furthermore, we developed a number of approximations to our model. While the deterministic approximation builds the intuition of offering margin ordered sets, the normal and integer programming approximations allow us to offer a single set over the whole selling horizon.

There are several interesting directions for future research. It would be useful to further study the version of the problem where only one set of products is offered over the whole selling horizon. We were fortunate in the sense that our normal and integer programming approximations ended up offering a single set, but one can a priori try to impose the requirement that one set of products should be offered in problem (1)-(3). If we need to offer a single set, then we can capture this situation by imposing the constraints $y(S) \in \{0, 1\}$ for all $S \subseteq \{1, \ldots, n\}$ in problem (1)-(3). Even when we have such binary constraints in problem (1)-(3), we can still formulate an equivalent nonlinear integer program to this problem simply by replacing the constraints $w_i/v_i - w_0/v_0 \leq 0$ for all $i = 1, \ldots, n$ in problem (4)-(7) with the constraints $w_i/v_i \in \{0, w_0/v_0\}$ for all $i = 1, \ldots, n$. By following precisely the same argument in the proof of Theorem 1, we can show that the version of problem (1)-(3) with binary constraints on the decision variables y is equivalent to the version of problem (4)-(7) with constraints (6) replaced with $w_i/v_i \in \{0, w_0/v_0\}$ for all $i = 1, \ldots, n$. However, it turns out that if we replace constraints (6) in problem (4)-(7) with $w_i/v_i \in \{0, w_0/v_0\}$ for all $i = 1, \ldots, n$, then the dynamic programming approach in Section 4 does not work any more. The difficulty is that for an arbitrary value of w_0, by choosing $w_i/v_i \in \{0, w_0/v_0\}$ for all $i = 1, \ldots, n$, we are not guaranteed to have $\sum_{i=1}^n w_i + w_0 = 1$. In fact, the only possible values of w_0 that allow us to choose $w_i/v_i \in \{0, w_0/v_0\}$ for all $i = 1, \ldots, n$ and still have $\sum_{i=1}^n w_i + w_0 = 1$ are those that satisfy $w_0 + \sum_{i \in S} v_i \, w_0/v_0 = 1$ for some $S \subseteq \{1, \ldots, n\}$. The computational effort for finding all such possible values of w_0 grows exponentially with n. Therefore, although our normal and integer programming approximations provide useful approximate methods, it is still difficult to accurately solve the variant of problem (1)-(3) with the constraint $y(S) \in \{0, 1\}$ for all $S \subseteq \{1, \ldots, n\}$. Another possible direction for future research is that we can look for ways of imposing additional constraints on offered sets. Finally, it is useful to see whether our results can be extended under customer choice models that are more complicated than the multinomial logit model.

Acknowledgements

The author thanks the senior editor and the two anonymous referees whose comments substantially improved the paper. He also acknowledges many stimulating conversations he has had with Guillermo Gallego and Paat Rusmevichientong on choice modeling. The research of the author was supported in part by the National Science Foundation grants CMMI 0825004 and CMMI 0969113.

References

Bront, J. J. M., Diaz, I. M. and Vulcano, G. (2009), 'A column generation algorithm for choice-based network revenue management', *Operations Research* **57**(3), 769–784.

Gallego, G., Iyengar, G., Phillips, R. and Dubey, A. (2004), Managing flexible products on a network, Computational Optimization Research Center Technical Report TR-2004-01, Columbia University.

Gallego, G. and Moon, I. K. (1993), 'The distribution-free newsboy problem - Review and extensions', *Journal of the Operational Research Society* **44**(8), 825–834.

Gallego, G., Ratliff, R. and Shebalov, S. (2011), A general attraction model and an efficient formulation for the network revenue management problem, Technical report, Columbia University, New York, NY.

Gallego, G. and van Ryzin, G. (1994), 'Optimal dynamic pricing of inventories with stochastic demand over finite horizons', *Management Science* **40**(8), 999–1020.

Gallego, G. and van Ryzin, G. (1997), 'A multiproduct dynamic pricing problem and its applications to yield management', *Operations Research* **45**(1), 24–41.

Gaur, V. and Honhon, D. (2006), 'Assortment planning and inventory decisions under a locational choice model', *Management Science* **52**(10), 1528–1543.

Goyal, V., Levi, R. and Segev, D. (2009), Near-optimal algorithms for the assortment planning problem under dynamic substitution and stochastic demand, Technical report, MIT, Cambridge, MA.

Honhon, D., Gaur, V. and Seshadri, S. (2010), 'Assortment planning and inventory decisions under stock-out based substitution', *Operations Research* **58**(5), 1364–1379.

Johnson, L. A. and Montgomery, D. C. (1974), *Operations Research in Production Planning, Scheduling and Inventory Control*, John Wiley & Sons, New York.

Kok, A. G. and Fisher, M. L. (2007), 'Demand estimation and assortment optimization under substitution: Methodology and application', *Operations Research* **55**(6), 1001–10021.

Lancaster, K. J. (1966), 'A new approach to consumer theory', *The Journal of Political Economy* **74**(2), 132–157.

Levi, R. and Radovanovic, A. (2010), 'Provably near-optimal LP-based policies for revenue management in systems with reusable resources', *Operations Research* **58**(2), 503–507.

Liu, Q. and van Ryzin, G. (2008), 'On the choice-based linear programming model for network revenue management', *Manufacturing & Service Operations Management* **10**(2), 288–310.

Mahajan, S. and van Ryzin, G. (2001), 'Stocking retail assortments under dynamic customer substitution', *Operations Research* **49**(3), 334–351.

McFadden, D. (1980), 'Econometric models for probabilistic choice among products', *The Journal of Business* **53**(3), S13–29.

Smith, S. and Agrawal, N. (2000), 'Management of multi-item retail inventory systems with demand substitution', *Operations Research* **48**(1), 50–64.

Talluri, K. T. and van Ryzin, G. J. (2005), *The Theory and Practice of Revenue Management*, Springer, New York, NY.

Talluri, K. and van Ryzin, G. (1998), 'An analysis of bid-price controls for network revenue management', *Management Science* **44**(11), 1577–1593.

Talluri, K. and van Ryzin, G. (2004), ‘Revenue management under a general discrete choice model of consumer behavior’, *Management Science* **50**(1), 15–33.

van Ryzin, G. and Mahajan, S. (1999), ‘On the relationship between inventory costs and variety benefits in retailassortments’, *Management Science* **45**(11), 1496–1509.

van Ryzin, G. and Vulcano, G. (2008), ‘Computing virtual nesting controls for network revenue management under customer choice behavior’, *Manufacturing & Service Operations Management* **10**(3), 448–467.

Zhang, D. and Cooper, W. L. (2005), ‘Revenue management for parallel flights with customer choice behavior’, *Operations Research* **53**(3), 415–431.

A Appendix: Omitted Proofs

A.1 Proof of Theorem 1

First, assuming that (y^*, x^*) is an optimal solution to problem (1)-(3) providing the objective value Z^*, we define the solution (w^*, w_0^*, x^*) to problem (4)-(7) as in (8) in the first part of the proof of Theorem 1 in the main text. By the definition of the multinomial logit model, we have $P_i(S)/v_i = 1/V(S)$ if $i \in S$ and $P_i(S)/v_i = 0$ otherwise. Similarly, we have $P_0(S)/v_0 = 1/V(S)$. In this case, letting $\mathbf{1}(\cdot)$ be the indicator function and using the definitions of w^* and w_0^* in (8), it follows that $w_i^*/v_i = \sum_{S\subseteq\{1,\ldots,n\}} P_i(S)\, y^*(S)/v_i = \sum_{S\subseteq\{1,\ldots,n\}} \mathbf{1}(i \in S)\, y^*(S)/V(S) \leq \sum_{S\subseteq\{1,\ldots,n\}} y^*(S)/V(S) = \sum_{S\subseteq\{1,\ldots,n\}} P_0(S)\, y^*(S)/v_0 = w_0^*/v_0$ for all $i = 1, \ldots, n$. Furthermore, we observe that

$$\sum_{i=1}^n w_i^* + w_0^* = \sum_{i=1}^n \sum_{S\subseteq\{1,\ldots,n\}} P_i(S)\, y^*(S) + \sum_{S\subseteq\{1,\ldots,n\}} P_0(S)\, y^*(S)$$
$$= \sum_{S\subseteq\{1,\ldots,n\}} \left[\sum_{i=1}^n P_i(S) + P_0(S)\right] y^*(S) = \sum_{S\subseteq\{1,\ldots,n\}} y^*(S) = 1,$$

where the third equality follows by noting that $\sum_{i=1}^n P_i(S) + P_0(S) = 1$ and the fourth equality follows from constraint (2). Therefore, the solution (w^*, w_0^*, x^*) is feasible to problem (4)-(7). Noting that we have $\lambda\, w_i^* = \sum_{S\subseteq\{1,\ldots,n\}} \lambda\, P_i(S)\, y^*(S)$ by (8), the objective value provided by the solution (w^*, w_0^*, x^*) for problem (4)-(7) is also Z^*.

Second, assuming that (w^*, w_0^*, x^*) is an optimal solution to problem (4)-(7) providing the objective value Z^*, we define the solution (y^*, x^*) to problem (1)-(3) as in (9) in the second part of the proof of Theorem 1 in the main text. Since the products are ordered so that $w_1^*/v_1 \geq w_2^*/v_2 \geq \ldots \geq w_n^*/v_n$ and we have $w_0^*/v_0 \geq w_1^*/v_1$ by constraints (6), the definition of y^* in (9) implies that $y^* \geq 0$. By the definition of y^*, only the decision variables $y^*(S_0), y^*(S_1), \ldots, y^*(S_n)$ among $\{y^*(S) : S \subseteq \{1, \ldots, n\}\}$ can possibly take nonzero values and we obtain

$$\sum_{S\subseteq\{1,\ldots,n\}} y^*(S) = \sum_{i=0}^n y^*(S_i) = \sum_{i=0}^{n-1} \left[\frac{w_i^*}{v_i} - \frac{w_{i+1}^*}{v_{i+1}}\right] V(S_i) + \frac{w_n^*}{v_n} V(S_n)$$
$$= \frac{w_0^*}{v_0} V(S_0) + \sum_{i=1}^n \frac{w_i^*}{v_i} [V(S_i) - V(S_{i-1})] = \frac{w_0^*}{v_0} v_0 + \sum_{i=1}^n \frac{w_i^*}{v_i} v_i = 1, \quad (24)$$

where the fourth equality uses the fact that $V(S_0) = v_0$ and $V(S_i) - V(S_{i-1}) = v_i$ by definition and the fifth equality follows from constraint (5). Therefore, the solution (y^*, x^*) is feasible to problem (1)-(3). To see that the solution (y^*, x^*) provides the objective value Z^* for problem (1)-(3), we note that only the sets $S_i, S_{i+1}, \ldots, S_n$ among $S_0, S_1, \ldots, S_n$ include product i so that we obtain

$$\sum_{S\subseteq\{1,\ldots,n\}} \lambda\, P_i(S)\, y^*(S) = \lambda\, P_i(S_i)\, y^*(S_i) + \lambda\, P_i(S_{i+1})\, y^*(S_{i+1}) + \ldots + \lambda\, P_i(S_n)\, y^*(S_n)$$
$$= \lambda\, v_i \left[\frac{w_i^*}{v_i} - \frac{w_{i+1}^*}{v_{i+1}}\right] + \lambda\, v_i \left[\frac{w_{i+1}^*}{v_{i+1}} - \frac{w_{i+2}^*}{v_{i+2}}\right] + \ldots + \lambda\, v_i \frac{w_n^*}{v_n} = \lambda\, w_i^*,$$

where the second equality follows from the definition of $y^*(S)$ in (9) and the fact that $v_i = P_i(S)\, V(S)$ for all $i \in S$ by the definition of $P_i(S)$ in the multinomial logit model. $\square$

A.2 Extreme Points of the Set of Feasible Solutions to Problem (4)-(7)

The next corollary is used in the discussion in Section 3 and in the proofs of Propositions 5 and 6.

Corollary 8 *For any $S \subseteq \{1, \ldots, n\}$, define the point $(\tilde{w}(S), \tilde{w}_0(S)) \in \Re^{n+1}$ such that $\tilde{w}_i(S) = P_i(S)$ for all $i = 0, \ldots, n$. Any extreme point of the set of feasible solutions to problem* (4)-(7) *is one of the points $\{(\tilde{w}(S), \tilde{w}_0(S)) : S \subseteq \{1, \ldots, n\}\}$.*

Proof. The ideas we use are very similar to those in the proof of Theorem 1. Choose any point (w', w'_0) that is in the set of feasible solutions to problem (4)-(7). We want to show that (w', w'_0) can be written as a convex combination of the points in $\{(\tilde{w}(S), \tilde{w}_0(S)) : S \subseteq \{1, \ldots, n\}\}$. We order the products so that $w'_1/v_1 \geq w'_2/v_2 \geq \ldots \geq w'_n/v_n$. Since (w', w'_0) is feasible to problem (4)-(7), constraints (6) also imply that $w'_0/v_0 \geq w'_1/v_1$. We label the sets $S_0 = \varnothing$ and $S_i = \{1, \ldots, i\}$ for all $i = 1, \ldots, n$. For all $S \subseteq \{1, \ldots, n\}$, we define $\alpha'(S)$ as follows. For all $i = 0, \ldots, n$, we let

$$\alpha'(S_i) = \left[\frac{w'_i}{v_i} - \frac{w'_{i+1}}{v_{i+1}}\right] V(S_i)$$

with the convention that $\alpha'(S_n) = [w'_n/v_n]\, V(S_n)$. For all other sets S, we let $\alpha'(S) = 0$. By using the same argument that we follow in (24), we have $\sum_{S \subseteq \{1,\ldots,n\}} \alpha'(S) = 1$ and $\alpha'(S) \geq 0$, in which case, $\{\alpha'(S) : S \subseteq \{1, \ldots, n\}\}$ are convex weights. For any $k = 0, \ldots, n$, we focus on the kth components of the points $\{(\tilde{w}(S), \tilde{w}_0(S)) : S \subseteq \{1, \ldots, n\}\}$ to write

$$\begin{aligned}\sum_{S \subseteq \{1,\ldots,n\}} \alpha'(S)\, \tilde{w}_k(S) = \sum_{i=0}^{n} \alpha'(S_i)\, \tilde{w}_k(S_i) &= \sum_{i=0}^{n} \left[\frac{w'_i}{v_i} - \frac{w'_{i+1}}{v_{i+1}}\right] V(S_i)\, P_k(S_i) \\ &= \sum_{i=0}^{n} \left[\frac{w'_i}{v_i} - \frac{w'_{i+1}}{v_{i+1}}\right] V(S_i) \frac{\mathbf{1}(k \in S_i)\, v_k}{V(S_i)} = \sum_{i=k}^{n} \left[\frac{w'_i}{v_i} - \frac{w'_{i+1}}{v_{i+1}}\right] v_k = w'_k,\end{aligned}$$

where the first equality uses the fact that $\alpha'(S) = 0$ whenever S is not one of the sets $S_0, \ldots, S_n$, the second equality is by noting that $\tilde{w}_k(S_i) = P_k(S_i)$ by definition, the third equality follows from the definition of $P_i(S)$ and the fourth equality is by noting that $k \in S_i$ only for $i = k, k+1, \ldots, n$. For any $k = 0, \ldots, n$, the chain of equalities above show that the kth component of (w', w'_0) can be written as a convex combination of the kth components of the points $\{(\tilde{w}(S), \tilde{w}_0(S)) : S \subseteq \{1, \ldots, n\}\}$ by using the convex weights $\{\alpha'(S) : S \subseteq \{1, \ldots, n\}\}$. Thus, (w', w'_0) can be written as a convex combination of the points in $\{(\tilde{w}(S), \tilde{w}_0(S)) : S \subseteq \{1, \ldots, n\}\}$ and we obtain the desired result. $\square$

A.3 Proof of Proposition 3

Fix w_i and let $\hat{x}_i$ be the smallest x_i that satisfies $\mathbb{P}\{\mathsf{Pois}(\lambda\, w_i) \leq x_i\} \geq 1 - c_i/p_i$. This is to say that $X_i(w_i) = \hat{x}_i$. In the first part of the proof, we assume that $\mathbb{P}\{\mathsf{Pois}(\lambda\, w_i) \leq \hat{x}_i\} > 1 - c_i/p_i$. Since the distribution function of a Poisson random variable is continuous in its mean, for $\epsilon > 0$ small enough, the smallest x_i that satisfies $\mathbb{P}\{\mathsf{Pois}(\lambda\, (w_i + \epsilon)) \leq x_i\} \geq 1 - c_i/p_i$ is also $\hat{x}_i$ so that $X_i(w_i + \epsilon) = \hat{x}_i$. In this case, using $a \wedge b$ to denote $\min\{a, b\}$, the definition of $G_i(\cdot)$ implies that

$$G_i(w_i + \epsilon) - G_i(w_i) = p_i \, \mathbb{E}\big\{(\mathsf{Pois}(\lambda\,(w_i + \epsilon)) \wedge \hat{x}_i) - (\mathsf{Pois}(\lambda\, w_i) \wedge \hat{x}_i)\big\}$$
$$= p_i \, \mathbb{E}\big\{[(\mathsf{Pois}(\lambda\, w_i) + \mathsf{Pois}(\lambda\,\epsilon)) \wedge \hat{x}_i] - (\mathsf{Pois}(\lambda\, w_i) \wedge \hat{x}_i)\big\}, \quad (25)$$

where $\mathsf{Pois}(\lambda\, w_i)$ and $\mathsf{Pois}(\lambda\,\epsilon)$ are independent of each other. For the Poisson random variable $\mathsf{Pois}(\lambda\,\epsilon)$, we have $\mathbb{P}\{\mathsf{Pois}(\lambda\,\epsilon) = 1\} = \lambda\,\epsilon + o(\epsilon)$, where $o(\epsilon)$ denotes any function $h(\cdot)$ that satisfies $\lim_{\epsilon \to 0} h(\epsilon)/\epsilon = 0$. Conditioning on $\mathsf{Pois}(\lambda\,\epsilon)$, the last expectation in (25) can be written as

$$\mathbb{E}\big\{[(\mathsf{Pois}(\lambda\, w_i) + \mathsf{Pois}(\lambda\,\epsilon)) \wedge \hat{x}_i] - (\mathsf{Pois}(\lambda\, w_i) \wedge \hat{x}_i)\big\}$$
$$= \mathbb{E}\big\{[(\mathsf{Pois}(\lambda\, w_i) + 1) \wedge \hat{x}_i] - (\mathsf{Pois}(\lambda\, w_i) \wedge \hat{x}_i) \,|\, \mathsf{Pois}(\lambda\,\epsilon) = 1\big\} \,(\lambda\,\epsilon + o(\epsilon))$$
$$+ \mathbb{E}\big\{[(\mathsf{Pois}(\lambda\, w_i) + \mathsf{Pois}(\lambda\,\epsilon)) \wedge \hat{x}_i] - (\mathsf{Pois}(\lambda\, w_i) \wedge \hat{x}_i) \,|\, \mathsf{Pois}(\lambda\,\epsilon) \geq 2\big\} \, \mathbb{P}\big\{\mathsf{Pois}(\lambda\,\epsilon) \geq 2\big\}. \quad (26)$$

Working with the first conditional expectation on the right side above, we observe that we can drop the condition since $\mathsf{Pois}(\lambda\, w_i)$ and $\mathsf{Pois}(\lambda\,\epsilon)$ are independent of each other. Furthermore, we have $[(\mathsf{Pois}(\lambda\, w_i) + 1) \wedge \hat{x}_i] - (\mathsf{Pois}(\lambda\, w_i) \wedge \hat{x}_i) = 1$ whenever $\mathsf{Pois}(\lambda\, w_i) + 1 \leq \hat{x}_i$. Otherwise, the last difference is equal to zero. Therefore, the first conditional expectation on the right side of (26) is equal to $\mathbb{P}\{\mathsf{Pois}(\lambda\, w_i) + 1 \leq \hat{x}_i\}$. Focusing on the second term on the right side of (26), we claim that this term is $o(\epsilon)$. To see this claim, note that $(a \wedge c) - (b \wedge c) \leq |a - b|$. Thus, we have

$$0 \leq \mathbb{E}\big\{[(\mathsf{Pois}(\lambda\, w_i) + \mathsf{Pois}(\lambda\,\epsilon)) \wedge \hat{x}_i] - (\mathsf{Pois}(\lambda\, w_i) \wedge \hat{x}_i) \,|\, \mathsf{Pois}(\lambda\,\epsilon) \geq 2\big\} \leq \mathbb{E}\big\{\mathsf{Pois}(\lambda\,\epsilon) \,|\, \mathsf{Pois}(\lambda\,\epsilon) \geq 2\big\},$$

which implies that the claim follows if we can show that $\mathbb{E}\{\mathsf{Pois}(\lambda\,\epsilon) \,|\, \mathsf{Pois}(\lambda\,\epsilon) \geq 2\} \, \mathbb{P}\{\mathsf{Pois}(\lambda\,\epsilon) \geq 2\} = o(\epsilon)$. Noting the identity, which can be obtained by conditioning on $\mathsf{Pois}(\lambda\,\epsilon)$,

$$\lambda\,\epsilon = \mathbb{E}\big\{\mathsf{Pois}(\lambda\,\epsilon)\big\} = e^{-\lambda\,\epsilon}\,\lambda\,\epsilon + \mathbb{E}\{\mathsf{Pois}(\lambda\,\epsilon) \,|\, \mathsf{Pois}(\lambda\,\epsilon) \geq 2\} \, \mathbb{P}\{\mathsf{Pois}(\lambda\,\epsilon) \geq 2\},$$

we have $\mathbb{E}\{\mathsf{Pois}(\lambda\,\epsilon) \,|\, \mathsf{Pois}(\lambda\,\epsilon) \geq 2\} \, \mathbb{P}\{\mathsf{Pois}(\lambda\,\epsilon) \geq 2\} = \lambda\epsilon\,(1 - e^{-\lambda\,\epsilon}) = o(\epsilon)$, where the second equality holds since $\lim_{\epsilon \to 0} \lambda\epsilon\,(1 - e^{-\lambda\,\epsilon})/\epsilon = 0$. This establishes the claim that the second term on the right side of (26) is $o(\epsilon)$. Since the first conditional expectation on the right side of (26) has been shown to be $\mathbb{P}\{\mathsf{Pois}(\lambda\, w_i) + 1 \leq \hat{x}_i\}$, (26) implies that $\mathbb{E}\big\{[(\mathsf{Pois}(\lambda\, w_i) + \mathsf{Pois}(\lambda\,\epsilon)) \wedge \hat{x}_i] - (\mathsf{Pois}(\lambda\, w_i) \wedge \hat{x}_i)\big\} = \mathbb{P}\{\mathsf{Pois}(\lambda\, w_i) + 1 \leq \hat{x}_i\} \,(\lambda\,\epsilon + o(\epsilon)) + o(\epsilon)$. Using this equality in (25), we have

$$\lim_{\epsilon \downarrow 0} \frac{G_i(w_i + \epsilon) - G_i(w_i)}{\epsilon} = p_i \lim_{\epsilon \downarrow 0} \frac{\mathbb{P}\{\mathsf{Pois}(\lambda\, w_i) + 1 \leq \hat{x}_i\} \,\lambda\,\epsilon + o(\epsilon)}{\epsilon}$$
$$= p_i \, \mathbb{P}\{\mathsf{Pois}(\lambda\, w_i) + 1 \leq \hat{x}_i\} \,\lambda \leq p_i (1 - c_i/p_i)\,\lambda = (p_i - c_i)\,\lambda,$$

where the inequality above follows by noting the fact that $\hat{x}_i$ is the smallest value of x_i that satisfies $\mathbb{P}\{\mathsf{Pois}(\lambda\, w_i) \leq x_i\} \geq 1 - c_i/p_i$ so that we must have $\mathbb{P}\{\mathsf{Pois}(\lambda\, w_i) \leq \hat{x}_i - 1\} < 1 - c_i/p_i$. Thus, the right derivative of $G_i(\cdot)$ at w_i is bounded by $(p_i - c_i)\,\lambda$. To bound the left derivative of $G_i(\cdot)$ at w_i, we write $G_i(w_i) - G_i(w_i - \epsilon) = p_i \, \mathbb{E}\big\{[(\mathsf{Pois}(\lambda\,(w_i - \epsilon)) + \mathsf{Pois}(\lambda\,\epsilon)) \wedge \hat{x}_i] - (\mathsf{Pois}(\lambda\,(w_i - \epsilon)) \wedge \hat{x}_i)\big\}$ and the same line of reasoning above shows that the left derivative of $G_i(\cdot)$ at w_i is bounded by $(p_i - c_i)\,\lambda$.

In the second part of the proof, we assume that $\mathbb{P}\{\mathsf{Pois}(\lambda\, w_i) \leq \hat{x}_i\} = 1 - c_i/p_i$. In this case, we have $\mathbb{P}\{\mathsf{Pois}(\lambda\,(w_i + \epsilon)) \leq \hat{x}_i\} < 1 - c_i/p_i$ so that for small enough $\epsilon > 0$, it must be the case that

$X_i(w_i+\epsilon)=\hat{x}_i+1$. Similar to (25), the definition of $G_i(\cdot)$ implies that

$$\begin{aligned} G_i(w_i+\epsilon)-G_i(w_i) &= p_i\,\mathbb{E}\big\{(\mathsf{Pois}(\lambda\,(w_i+\epsilon))\wedge(\hat{x}_i+1))-(\mathsf{Pois}(\lambda\,w_i)\wedge\hat{x}_i)\big\}-c_i \\ &= p_i\,\mathbb{E}\big\{[(\mathsf{Pois}(\lambda\,w_i)+\mathsf{Pois}(\lambda\,\epsilon))\wedge(\hat{x}_i+1)]-(\mathsf{Pois}(\lambda\,w_i)\wedge\hat{x}_i)\big\}-c_i. \end{aligned} \tag{27}$$

For the Poisson random variable $\mathsf{Pois}(\lambda\,\epsilon)$, we have $\mathbb{P}\{\mathsf{Pois}(\lambda\,\epsilon)=0\}=1-\lambda\,\epsilon+o(\epsilon)$. Conditioning on $\mathsf{Pois}(\lambda\,\epsilon)$, we write the expectation on the right side above as

$$\begin{aligned} &\mathbb{E}\big\{[(\mathsf{Pois}(\lambda\,w_i)+\mathsf{Pois}(\lambda\,\epsilon))\wedge(\hat{x}_i+1)]-(\mathsf{Pois}(\lambda\,w_i)\wedge\hat{x}_i)\big\} \\ &\quad=\mathbb{E}\big\{[\mathsf{Pois}(\lambda\,w_i)\wedge(\hat{x}_i+1)]-(\mathsf{Pois}(\lambda\,w_i)\wedge\hat{x}_i)\big\}\,(1-\lambda\,\epsilon+o(\epsilon)) \\ &\qquad+\mathbb{E}\big\{[(\mathsf{Pois}(\lambda\,w_i)+1)\wedge(\hat{x}_i+1)]-(\mathsf{Pois}(\lambda\,w_i)\wedge\hat{x}_i)\big\}\,(\lambda\,\epsilon+o(\epsilon)) \\ &\qquad\;+\mathbb{E}\big\{[(\mathsf{Pois}(\lambda\,w_i)+\mathsf{Pois}(\lambda\,\epsilon))\wedge(\hat{x}_i+1)]-(\mathsf{Pois}(\lambda\,w_i)\wedge\hat{x}_i)\,|\,\mathsf{Pois}(\lambda\,\epsilon)\geq 2\big\}\,\mathbb{P}\big\{\mathsf{Pois}(\lambda\,\epsilon)\geq 2\big\}. \end{aligned} \tag{28}$$

For the first expectation on the right side of (28), we have $[\mathsf{Pois}(\lambda\,w_i)\wedge(\hat{x}_i+1)]-(\mathsf{Pois}(\lambda\,w_i)\wedge\hat{x}_i)=1$ whenever $\mathsf{Pois}(\lambda\,w_i)\geq\hat{x}_i+1$. Otherwise, the last difference is zero. Thus, the first expectation on the right side of (28) is $\mathbb{P}\{\mathsf{Pois}(\lambda\,w_i)\geq\hat{x}_i+1\}$. For the second expectation on the right side of (28), we have $[(\mathsf{Pois}(\lambda\,w_i)+1)\wedge(\hat{x}_i+1)]-(\mathsf{Pois}(\lambda\,w_i)\wedge\hat{x}_i)=1$, so that the second expectation is one. Using the same line of reasoning in the first part of the proof, it is possible to show that the third term on the right side of (28) is $o(\epsilon)$. If we use these observations in (28), then (27) implies that

$$\begin{aligned} G_i(w_i+\epsilon)-G_i(w_i) &= p_i\,\mathbb{P}\{\mathsf{Pois}(\lambda\,w_i)\geq\hat{x}_i+1\}\,(1-\lambda\,\epsilon+o(\epsilon))+p_i\,(\lambda\,\epsilon+o(\epsilon))+p_i\,o(\epsilon)-c_i \\ &= p_i\,(c_i/p_i)\,(1-\lambda\,\epsilon+o(\epsilon))+p_i\,(\lambda\,\epsilon+o(\epsilon))+p_i\,o(\epsilon)-c_i=(p_i-c_i)\,\lambda\,\epsilon+o(\epsilon), \end{aligned}$$

where the second equality is by the fact that $\mathbb{P}\{\mathsf{Pois}(\lambda\,w_i)\leq\hat{x}_i\}=1-c_i/p_i$. Dividing the chain of equalities above by ϵ and taking the limit as ϵ goes to zero, we see that the right derivative of $G_i(\cdot)$ at w_i is exactly $(p_i-c_i)\,\lambda$. By following the same argument, it is possible to check that the left derivative of $G_i(\cdot)$ at w_i is bounded by $(p_i-c_i)\,\lambda$ and we obtain the desired result. □

A.4 Proof of Proposition 4

Noting that $\min\{a,b\}$ is a concave function of a, Jensen's inequality implies that $\sum_{i=1}^n p_i\,\min\{\lambda\,w_i,x_i\}-\sum_{i=1}^n c_i\,x_i\;\geq\;\sum_{i=1}^n p_i\,\mathbb{E}\{\min\{\mathsf{Pois}(\lambda\,w_i),x_i\}\}-\sum_{i=1}^n c_i\,x_i$. Since problem (15)-(18) is equivalent to maximizing $\sum_{i=1}^n p_i\,\min\{\lambda\,w_i,x_i\}-\sum_{i=1}^n c_i\,x_i$ over $(w,w_0,x)\geq 0$ subject to constraints (5) and (6), the optimal objective value of problem (15)-(18) is an upper bound on the optimal objective value of problem (4)-(7). Thus, using Z^D to denote the optimal objective value of problem (15)-(18) and noting that $\sum_{i=1}^n Q_i(w_i^*,X_i(w_i^*))$ is the optimal objective value of problem (4)-(7), we have $Z^D\geq\sum_{i=1}^n Q_i(w_i^*,X_i(w_i^*))$. On the other hand, $X_i(w_i)$ is defined as the maximizer of $Q_i(w_i,x_i)=p_i\,\mathbb{E}\{\min\{\mathsf{Pois}(\lambda\,w_i),x_i\}\}-c_i\,x_i$ over x_i. Thus, $Q_i(w_i,X_i(w_i))\geq Q_i(w_i,x_i)$ for any stocking quantity x_i. These observations imply that

$$\frac{\sum_{i=1}^n Q_i(w_i^D,X_i(w_i^D))}{\sum_{i=1}^n Q_i(w_i^*,X_i(w_i^*))}\geq\frac{\sum_{i=1}^n Q_i(w_i^D,X_i(w_i^D))}{Z^D}\geq\frac{\sum_{i=1}^n Q_i(w_i^D,\lambda\,w_i^D)}{Z^D},$$

in which case, the result follows if we can show that the expression on the right side of the chain of inequalities above upper bounds the expression on the right side of (19). We have

$$\begin{aligned}\frac{\sum_{i=1}^n Q_i(w_i^D, \lambda\, w_i^D)}{Z^D} &= \frac{\sum_{i=1}^n p_i\, \mathbb{E}\big\{\min\{\mathsf{Pois}(\lambda\, w_i^D), \lambda\, w_i^D\}\big\} - \sum_{i=1}^n c_i\, \lambda\, w_i^D}{Z^D} \\ &= \frac{\sum_{i=1}^n p_i\, \lambda\, w_i^D - \sum_{i=1}^n p_i\, \mathbb{E}\big\{\max\{\lambda\, w_i^D - \mathsf{Pois}(\lambda\, w_i^D), 0\}\big\} - \sum_{i=1}^n c_i\, \lambda\, w_i^D}{Z^D} \\ &= \frac{Z^D - \sum_{i=1}^n p_i\, \mathbb{E}\big\{\max\{\lambda\, w_i^D - \mathsf{Pois}(\lambda\, w_i^D), 0\}\big\}}{Z^D} \geq 1 - \frac{\sum_{i=1}^n p_i \sqrt{\lambda\, w_i^D}}{2\, Z^D},\end{aligned}$$

where the third equality uses the fact that (w^D, w_0^D) is an optimal solution to problem (15)-(18) and the inequality follows since as shown by Gallego and Moon (1993), we have $\mathbb{E}\{[W]^+\} \leq \sigma/2$ for a random variable W with mean zero and standard deviation σ.

We have $\sum_{i=1}^n w_i^D \leq 1$. A simple exercise in nonlinear programming shows that the maximum of $\sum_{i=1}^n p_i \sqrt{w_i}$ subject to the constraints $\sum_{i=1}^n w_i \leq 1$ and $w \geq 0$ is $\sqrt{\sum_{i=1}^n p_i^2}$. Thus, $\sqrt{\sum_{i=1}^n p_i^2\, \lambda}$ upper bounds the numerator of the last fraction above. To get a crude lower bound on Z^D that appears in the denominator of the same fraction, we let $w_1 = v_1/(v_0 + v_1)$, $w_0 = v_0/(v_0 + v_1)$ and set all of the other decision variables to zero to get a feasible solution to problem (15)-(18). In this case, Z^D is lower bounded by $(p_1 - c_1)\, \lambda\, \frac{v_1}{v_0+v_1}$. Using these bounds on the numerator and denominator of the last fraction above, it follows that

$$\frac{\sum_{i=1}^n Q_i(w_i^D, \lambda\, w_i^D)}{Z^D} \geq 1 - \frac{\sum_{i=1}^n p_i \sqrt{\lambda\, w_i^D}}{2\, Z^D} \geq 1 - \frac{\sqrt{\sum_{i=1}^n p_i^2\, \lambda}}{2\, (p_1 - c_1)\, \lambda\, \frac{v_1}{v_0+v_1}}. \qquad \square$$